ADVANCES IN MATHEMATICS 25, 59–90 (1977)

Decidability Criteria for the Similarity Problem, with Applications to the Moduli of Linear Dynamical Systems

CHRISTOPHER I. BYRNES*,[1] AND MICHAEL A. GAUGER†

Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01002

INTRODUCTION

The recognition of the importance of the problem of determining the similarity of matrices can be traced through more than a century of mathematics. The most widely known solutions depend on the existence of certain canonical forms. Triangularization can be traced to Jacobi's work in 1857 [13]. Jordan (1870) first introduced the canonical form bearing his name in working over finite fields [14], and Dickson (1902) extended the results to more general fields [6]. In fact, an algorithm can be described for obtaining the Jordan form of a triangular matrix or one whose eigenvalues are given. The rational canonical form was introduced by Frobenius (1879) for the complex field [10], and later improvements were made by Lattes (1914) [20], Kowalewski (1916)[19], and Dickson (1926) [7]. We recommend Aitken–Turnbull [1] and MacDuffee [21] for their treatments of these canonical forms, and especially for their historical notes, from which one can locate many of the early papers on matrices.

As for the barehanded calculation of the rational or Jordan form of an arbitrary matrix, only the former can be described by a practical finite algorithm. Such an algorithm for the latter would result in a method for solving polynomial equations in one variable—well known to be impossible for degree $\geqslant 5$. To write the rational form of a matrix A one calculates the invariant factors of its characteristic matrix $\lambda I - A$ by a cumbersome, yet effective, algorithm due to Smith and Frobenius [1, pp. 23–29, 49–56].

Motivated by the geometric invariant theory we proceed here to prove

THEOREM 3.6. *Let A and B be $n \times n$ matrices over a field* $\mathbf{k}$ *of characteristic zero. Then A is similar to B if and only if*:

(i) *A and B have the same characteristic polynomial, and*

* Partially supported by the NSF under Grant GJ-35759.
† Partially supported by the Grant GP-43814.
[1] Current address: Department of Mathematics, University of Utah, Salt Lake City, Utah 84112.

ISSN 0001-8708

(ii) *for all even* $k \leqslant 2n - 2$

$$\begin{aligned}\operatorname{rank}(A \otimes 1 - 1 \otimes A)^k &= \operatorname{rank}(B \otimes 1 - 1 \otimes B)^k \\ &= \operatorname{rank}(A \otimes 1 - 1 \otimes B)^k.\end{aligned}$$

There are philosophical as well as mathematical distinctions between this solution to the similarity problem and the solutions mentioned above. Given two matrices A and B, the classical determination of similarity is to calculate individually the canonical forms (invariant factors) and to compare them. Our method has as an essential ingredient certain algebraic calculations which simultaneously employ the entries of A and B. It does not give the similarity invariants of a single matrix.

The mathematical distinction can best be described by saying that our solution is "algebraic." The similarity of A and B can be decided by a finite number of polynomial equalities and inequalities in the entries of A and B. To be more precise, identify $M(n, \mathbf{k})$ with $\mathbf{k}^{n^2}$. Then, in the language of algebraic geometry, the set of all pairs (A, B) where A and B are similar is a constructible subset of $\mathbf{k}^{n^2} \times \mathbf{k}^{n^2}$. To say it is constructible is to say that it is a finite union of locally closed sets. A locally closed set can typically be described, in terms of global (polynomial) functions $f_1, \ldots, f_r, g_1, \ldots, g_s$ on $\mathbf{k}^{n^2} \times \mathbf{k}^{n^2}$, as being those common zeros of the f's for which at least one of the g's is nonvanishing. If $\phi_j(x)$ is the global function on $\mathbf{k}^{n^2}$ giving the jth coefficient of the characteristic polynomial of x, some of the $f(x, y)$'s are given by $\phi_j(x) - \phi_j(y)$. The rest of the f's and all of the g's arise from the rank computations in (ii) as minor determinant functions.

The process of obtaining a Jordan canonical form is not "algebraic" in that it involves the "irrational" and generally impossible problem of solving a 1-variable polynomial equation. One cannot express the eigenvalues of a matrix as global algebraic functions of its entries. On the other hand, the process of obtaining the rational form, or what is equivalent, its invariant factors, is algorithmic, although the mapping associating to a matrix its list of invariant factors cannot be algebraic. If it were, similarity classes would necessarily be Zariski-closed in $\mathbf{k}^{n^2}$, which is for the most part untrue (see Sect. 2).

In Section 2 we analyze the geometry of the similarity problem, in the context of an orbit space problem. This serves two immediate purposes. The first is to set notation and conventions for Sections 3, 4 and 5, and to provide geometric insight and motivation for the Main Theorem. The second is to work out, in a familiar example, some of the more basic notions (e.g., stability of orbits) involved in the general orbit space problem, thereby giving an exposition of these ideas intended for a general audience. These concepts also play an important role in Section 4, so that the results in Section 2 serve to contrast the geometry of two related algebraic problems. Section 3 is principally concerned with the proof of the Main Theorem, and includes a discussion of the existence of such decidability criteria for more general problems. In Section 4 we construct the moduli space for finite-dimensional, time-invariant, discrete-time linear dynamical systems, with a scalar input. This uses only affine invariant theory and the results in Section 2. We also indicate the structure of the moduli space for such systems with more general input, and identify some system-theoretic properties which are "generic," giving answers to some questions asked in [17]. In Section 5 we produce a canonical form (Theorem 5.2) for scalar input systems (i.e., pairs $(A, v) \in M(h, \mathbf{k}) \times \mathbf{k}^n$), and, using the Main Theorem, we establish decidability criteria (Theorem 5.3) for their equivalence. In the final section, we give decidability criteria for the conjugacy problem in the classical groups and their corresponding Lie algebras.

2. Similarity as an Orbit Space Problem

The more recent developments in the theory of canonical forms for $M(n, \mathbf{k})$ have been in the qualitative direction. In fact, our computational criterion has its roots in the rapproachment of algebraic geometry and linear control theory, viz., the global analysis of linear dynamical systems. Briefly, Mumford and Suominen have given [24, Chap. 1] an exposition of the similarity problem in terms of the representability of the moduli functor. Now, essentially contained in these results, which are for the most part negative, is the construction of a geometric quotient of a special class of linear dynamical systems by the relation of systems—theoretic equivalence (see Sect. 4). We should add that further computation yields a simple geometric proof of the canonical (or minimal) realization for such systems. The fine moduli space in [24] arises in the construction of a geometric quotient for the regular points of the similarity action of $G\ell(n, \mathbf{k})$ on $M(n, \mathbf{k})$, while an explicit analysis of the semistable and stable points hints at a reduction of the problem to simple cases.

We work only in characteristic 0 and over a fixed algebraically closed field $\mathbf{k}$, since similarity is well known to be an absolute property (i.e., similarity of matrices is not relative to the choice of a field extension $\mathbf{K}/\mathbf{k}$). Here is the geometric situation: Let $\mathbf{A}^{n^2}$ denote $\mathbf{k}^{n^2}$ with the Zariski topology. Thus if $f \in \mathbf{k}[x_{ij}]$, then the colocus of f,

$$\mathbf{A}_f^{n^2} = \{x = (x_{ij}) : f(x) \neq 0\},$$

is open, and is referred to as a principal affine open. For $x = (x_{ij}) \in \mathbf{A}^{n^2}$, the set

$$\mathcal{O}(x) = \{gxg^{-1} : g \in G\ell(n, \mathbf{k})\}$$

is referred to as the orbit of x, so that $\mathcal{O}(x) = \mathcal{O}(y)$ if and only if x and y are similar. Now, one would like to be able to parameterize the orbits of $G\ell(n, \mathbf{k})$ algebraically, so as to be able to speak of algebraic deformations or perturbations of matrices and their similarity classes. Thus, we ask for a variety structure on the set of orbits $\mathbf{A}^{n^2}/G\ell(n, \mathbf{k})$, related to $\mathbf{A}^{n^2}$ by

$$\mathbf{A}^{n^2} \xrightarrow{\pi} \mathbf{A}^{n^2}/G\ell(n, \mathbf{k}),$$

where the natural map π is algebraic. In this way, π would provide a parameterization for the similarity classes. Formally, a pair (V, ϕ) is a quotient for this action if $\mathbf{A}^{n^2} \to^{\phi} V$ is a categorical quotient (say, in the category of quasi-projective varieties). Clearly, in order that $\mathbf{A}^{n^2}/G\ell(n, \mathbf{k})$ carry such a structure, $\pi^{-1}([x]) = O(x)$ must be closed, for each x. This is where things begin to go awry.

By the closed orbit lemma (see [3, p. 98]) each orbit is a locally closed subvariety of $\mathbf{A}^{n^2}$ In fact, $\mathcal{O}(x)$ is open in $\mathscr{C}\ell(\mathcal{O}(x))$, so that $\mathscr{C}\ell(\mathcal{O}(x)) - \mathcal{O}(x)$ is a $G\ell(n, \mathbf{k})$-invariant set of dimension less than $\dim \mathcal{O}(x)$; i.e., $\mathscr{C}\ell(\mathcal{O}(x)) - \mathcal{O}(x)$ is the union of lower-dimensional orbits. In particular, orbits of minimal dimension provide examples of closed orbits. However in this case the orbits of minimal dimension are just the fixed points, which by Schur's lemma are precisely the scalar matrices. As this pathology suggests, there exist (in great profusion) orbits which are not closed. Such orbits may be easily constructed by choosing the appropriate Jordan block structure. More precisely, let $(x)_s$ denote the diagonal part of the Jordan canonical form of x.

PROPOSITION 2.1. *If $x_1, x_2 \in \mathbf{A}^{n^2}$, then $\mathscr{C}\ell(\mathcal{O}(x_1)) \cap \mathscr{C}\ell(\mathcal{O}(x_2)) \neq \phi$ if and only if $\mathcal{O}((x_1)_s) = \mathcal{O}((x_2)_s)$. In any case, the semisimple part of x_i, $(x_i)_s$, lies in $\mathscr{C}\ell(\mathcal{O}(x_i))$.*

See [24, p. 179], for a statement and proof of this proposition and for a well-known example illustrating the connection between the closedness of an orbit and Jordan block structure. By the closed orbit lemma this is also related to the dimension of an orbit and, in fact, we give a formula for $\dim \mathcal{O}(x)$ in terms of the block structure. First, notice

COROLLARY 2.2. *$\mathcal{O}(x)$ is closed if and only if $(x)_s \in \mathcal{O}(x)$; i.e., if and only if x is semisimple.*

In order to get at the dimension formula for orbits, we need to set some notation. For any $\lambda \in \mathbf{k}$, let $J(\lambda; m)$ be the $m \times m$ Jordan block with λ's on the diagonal, 1's on the superdiagonal, and 0's elsewhere. Also, if A and B are square matrices, $A \oplus B$ denotes the diagonal direct sum of the two block matrices.

PROPOSITION 2.3. *Suppose $x = \oplus_{i=1}^{s} (\oplus_{j=1}^{r_i} J(\lambda_i; n_{ij}))$, where $n_{i1} \leqslant \cdots \leqslant n_{ir_i}$, for all i. Then*

$$\dim \mathcal{O}(x) = n^2 - \sum_{i=1}^{s} \left(\sum_{j=1}^{r_i} (2r_i + 1 - 2j) n_{ij} \right). \tag{2.1}$$

Proof. Now $\alpha_x : G\ell(n, k) \to \mathcal{O}(x)$ is a rational surjection, where $\alpha_x(g) = gxg^{-1}$. Thus, $\dim \mathcal{O}(x) = \operatorname{rank} d\alpha_x(I)$, the derivative of α_x at the identity I in $G\ell(n, k)$. Since

$$d\alpha_x(I)(y) = \operatorname{ad}(x)(y) = [x, y],$$

$\dim \mathcal{O}(x) = n^2 - \dim Z(x)$, where $Z(x)$ denotes the centralizer of x in $M(n, \mathbf{k})$. Finally, the equality

$$\dim Z(x) = \sum_{i=1}^{s} \left(\sum_{j=1}^{r_i} (2r_i + 1 - 2j) n_{ij} \right)$$

is a straightforward matrix argument (see [11, pp. 146–147]). Q.E.D.

COROLLARY 2.4. *If x is semisimple, then*

$$\dim \mathcal{O}(x) = n^2 - \sum_{i=1}^{s} r_i^2, \tag{2.2}$$

where $1 \leqslant s \leqslant n$ is the number of distinct eigenvalues of x and r_i is the geometric multiplicity of the ith *eigenvalue.*

By Corollary 2.4, there exist closed orbits which are not of minimal dimension and, of course, these correspond to the cases $s > 1$ in formula (2.2). On the other hand, formula (2.2) achieves its maximum, $n^2 - n$, when $s = n$. In this case, all eigenvalues have multiplicity one, so that the characteristic and minimum polynomials are equal. Matrices with this latter property are classically referred to as *nonderogatory* matrices and, as is well known, these are precisely those matrices possessing a cyclic vector. The following theorem asserts that the set of nonderogatory matrices plays the important role of a principal orbit type; i.e., the union of those orbits having maximal dimension. As a motivation for this, consider the classical theorem: $\mathbf{k}[x] = Z(x)$ if and only if x is nonderogatory. Hence, by the Cayley–Hamilton theorem, $\dim Z(x)$ achieves its minimum value, n, on this set.

PROPOSITION 2.5. *x is nonderogatory if and only if* $\dim \mathcal{O}(x)$ *is maximal; i.e., if and only if* $\dim \mathcal{O}(x) = n^2 - n$. *Equivalently, the set of nonderogatory matrices is the set of regular points, denoted by* $(\mathbf{A}^{n^2})_{\text{reg}}$, *for the similarity action of* $G\ell(n, \mathbf{k})$ *and, thus, is Zariski open.*

Proof. To say that x is nonderogatory is to say that $r_i = 1$, for each $i = 1, \ldots, s$, and so n_{i1} is the geometric multiplicity of λ_i, for each i. Hence $\dim \mathcal{O}(x) = n^2 - \sum_{i=1}^{s} n_{i1} = n^2 - n$. It is clear that formula (2.1) takes on its maximum value, $n^2 - n$, when $r_i = 1$, for each i. Thus, the converse of the first statement also holds. By definition, x is regular for the similarity action of

$G\ell(n, \mathbf{k})$, since the stabilizer $S(x)$ has minimal dimension. Finally, the set of nonderogatory matrices, $(\mathbf{A}^{n^2})_{\mathrm{reg}}$, is open since $\dim \mathcal{O}(y)$ is an upper semicontinuous function of y (see [23, p. 10]). Q.E.D.

From the existence of nonclosed orbits, it is clear that $\mathbf{A}^{n^2}/G\ell(n, \mathbf{k})$ cannot be represented as a quotient variety of $\mathbf{A}^{n^2}$ by $G\ell(n, \mathbf{k})$. However, a categorical quotient still exists, by Mumford's theorem (see [9, p. 159]). We can get at this quotient as follows. In the coordinate ring $\mathbf{k}[x_{ij}]$ of $\mathbf{A}^{n^2}$, consider the subring of invariant polynomials; i.e., those $f \in \mathbf{k}[x_{ij}]$ such that $f^g(x) = f(gxg^{-1}) = f(x)$, for all $x \in \mathbf{A}^{n^2}$, for all $g \in G\ell(n, \mathbf{k})$. If $\mathbf{A}^{n^2} \xrightarrow{\phi} V$ is a categorical quotient, then any invariant f descends to a polynomial $f_1 \in \mathbf{k}[V]$ such that $f_1 \circ \phi = f$. Since any $g \in \mathbf{k}[V]$ can be lifted to an invariant, $g \circ \phi$, on $\mathbf{A}^{n^2}$, the coordinate ring $\mathbf{k}[V]$ of V must be the ring of invariants $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$. By the Hilbert–Nagata theorem (see [8, p. 42]), $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$ is finitely generated over $\mathbf{k}$ and, in fact, it is well known that the coefficients ϕ_ℓ of the characteristic polynomial

$$\phi(x, \lambda) = \sum_{\ell=1}^{n} \phi_\ell(x)\lambda^{n-\ell} + \lambda^n$$

generate $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$. Now the inclusion, $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})} \to \mathbf{k}[x_{ij}]$, induces a map

$$\mathrm{Hom}_{\mathbf{k}}(\mathbf{k}[x_{ij}], \mathbf{k}) \to \mathrm{Hom}_{\mathbf{k}}(\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}, \mathbf{k})$$

which is surjective (see [8, pp. 51–53]). We claim that the ϕ_ℓ's are algebraically independent. For if there were a dependence relation $P(\phi_\ell(x_{ij})) = 0$, then this relation would hold when evaluated on the diagonal matrices alone. In this case, $x_{ij} = \delta_{ij}\lambda_i$, so that one has $P(\phi_\ell(\lambda_i)) = 0$, where $\lambda_i \in \mathbf{k}$ are arbitrary. But this implies $P \equiv 0$, since the ϕ_ℓ are just the elementary symmetric functions of the λ_i's, which are well known to be algebraically independent (see [9, pp. 36–38]). Thus, by the Hilbert Nullstellensatz,

$$\mathbf{k}^{n^2} = \mathrm{Hom}_{\mathbf{k}}(\mathbf{k}[x_{ij}], \mathbf{k}) \qquad \text{and} \qquad \mathbf{k}^n = \mathrm{Hom}_{\mathbf{k}}(\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}, \mathbf{k}),$$

so that $\mathbf{A}^n$ is a concrete affine model for $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$:

PROPOSITION 2.6. *$\mathbf{A}^{n^2} \xrightarrow{\phi} \mathbf{A}^n$ is the categorical quotient of $\mathbf{A}^{n^2}$ by $G\ell(n, \mathbf{k})$, where $\phi(x) = (\phi_1(x),\ldots, \phi_n(x))$.*

Since any fiber of ϕ is the union of all orbits $\mathcal{O}(x)$ having the same "semisimple parts," the existence of nonclosed orbits is the major obstruction to the separation of similarity classes by regular invariants.

The first of two natural approaches to an algebraic determination of similarity is an analysis of the fiber of ϕ. By the existence of the Jordan canonical form, $\phi^{-1}(\phi(x))$ can contain only finitely many orbits. Moreover, the fiber is naturally a partially ordered set (ordered by the relation $\mathcal{O}(x) \leqslant \mathcal{O}(y)$ if and only if $\mathcal{O}(x) \subset \mathcal{C}\ell(\mathcal{O}(y))$), with a least element $0_{\phi(x)}$ and a greatest element $1_{\phi(x)}$. In this language, $0_{\phi(x)} = \mathcal{O}(y)$ where y is semisimple with $\phi(x) = \phi(y)$. In fact, $0_{\phi(x)}$ is the unique closed orbit in $\phi^{-1}(\phi(x))$ by Corollary 2.2, $1_{\phi(x)} = \mathcal{O}(y)$, where y is any nonderogatory matrix with $\phi(x) = \phi(y)$. By Proposition 2.5, $1_{\phi(x)}$ is a relatively open, everywhere dense subset of the irreducible affine variety $\phi^{-1}(\phi(x))$. Now there are two classical techniques for separating the orbits lying in a given fiber. The first assigns to any orbit its Weyr invariants (see [27]). Explicitly, if λ_i is an eigenvalue of y with multiplicity m_i, then, for $1 \leqslant j \leqslant m_i$, the ijth Weyr invariant, α_{ij}, of y is the rank of $(\lambda_i I - y)^j$. It is classical that $\{\phi_1(y), \ldots, \phi_n(y), \alpha_{ij}\}$ form a complete set of similarity invariants (see [21, pp. 73–74]). In particular, if y is nilpotent, then the integers rank y^j, for $1 \leqslant j \leqslant n$, form a complete set of similarity invariants (see Sect. 3, Proposition 3.3). However, this technique relies on a factorization of the characteristic polynomial into primes so that, although the Weyr invariants are effectively computable, if y has two distinct eigenvalues this computation is decidedly nonalgebraic (say, for $n \geqslant 5$). The second technique is entirely algorithmic and depends on significantly less information. One simply puts $\lambda I - y$ in its Smith normal form, thereby recovering the invariant factors (see [1, pp. 54–56]). As is well known, the characteristic polynomial and the invariant factors form a complete set of invariants for $\mathcal{O}(y)$. Although this algorithm is computable in a realistic sense, as we have pointed out in the introduction, it is not of an algebraic nature.

Remarks. 1. Notice that by Corollary 2.4 and Proposition 2.5, $0_{\phi(x)}$ and $1_{\phi(x)}$ are completely determined (within $\phi^{-1}(\phi(x))$ by dimension, so one may wonder whether dimension separates the intermediate orbits as well. Unfortunately, for $n = 4$ it may be directly verified that dim x = dim y = 10 for $\lambda_1 \neq \lambda_2$ and

$$x = \left(\bigoplus^{2} J(\lambda_1; 1)\right) \oplus J(\lambda_2; 2), \qquad y = J(\lambda_1; 2) \oplus \left(\bigoplus^{2} J(\lambda_2; 1)\right).$$

Thus the dimension formula cannot distinguish to which eigenvalues the blocks belong, despite the dependence of this formula on the block structure. It is exactly this flaw which forces us to consider the "two-variable" case; i.e., dimension theoretic computations for $\mathbf{A}^{n^2} \times_{\mathbf{A}^n} \mathbf{A}^{n^2}$.

2. Based on complete calculations for $n \leqslant 5$, and the one eigenvalue case for $n \leqslant 9$, we have found that the integers $\dim \mathcal{O}(x) - \dim \mathcal{O}(y)$ and $\dim \mathcal{O}(x \oplus x) - \dim \mathcal{O}(x \oplus y)$ separate distinct orbits in any fiber. In general (see Theorem 3.6), it appears unlikely that these integers should work as well as in the lower dimensions. However, it would be of practical, computational value to know the range in which they are complete invariants.

The intuition for the rational data listed in Theorem 3.6 also comes from a global study of the quotient $\mathbf{A}^{n^2} \to^{\phi} \mathbf{A}^n$; i.e., one may ask for modified fundamental domains. For example, let $n = 2$ and consider the principal affine open $\mathbf{A}_f^2$, where $f(x, y) = x^2 - 4y$. The map $\phi\colon \mathbf{A}^4_{f\circ\phi} \to \mathbf{A}_f^2$ is a good (more precisely, a geometric [23] or strict [9]) quotient of $\mathbf{A}^4_{f\circ\phi} = \phi^{-1}\mathbf{A}_f^2$ by $G\ell(2, \mathbf{k})$, since the fibers consist of one orbit, each fiber being an irreducible variety of dimension 2. Explicitly, $\phi(z) = (-\operatorname{tr} z, \det z)$ so that $f(\phi(z)) = 0$ if and only if z has a repeated eigenvalue. Thus, $\mathbf{A}^4_{f\circ\phi}$ consists of the maximal dimension, closed orbits in $\mathbf{A}^4$. In particular, the orbits for matrices $z \in \phi^{-1}(\mathbf{A}_f^2)$ are completely determined by $\phi(z)$ and by the nonvanishing of the polynomial defining $\mathbf{A}^4 - \phi^{-1}(\mathbf{A}_f^2)$; viz., the discriminant of the characteristic polynomial. Now it is possible to considerably enlarge the domain of ϕ, while retaining a geometric quotient. As a consequence of the identity $(\mathbf{A}^4)_{\text{reg}} \cap \phi^{-1}(\phi(x)) = \mathbf{1}_{\phi(x)}$, for all $x \in \mathbf{A}^4$, one may fill in the locus of f using the nonderogatory matrices with a repeated eigenvalue; i.e., $(\mathbf{A}^4)_{\text{reg}} \xrightarrow{\phi} \mathbf{A}^2$ is a geometric quotient. Thus, for $y \in (\mathbf{A}^4)_{\text{reg}}$, $\mathcal{O}(y)$ is completely determined by $\phi(y)$ and the nonvanishing of the polynomials defining $\mathbf{A}^4 - (\mathbf{A}^4)_{\text{reg}}$. It is easy to see that induction finishes the problem; i.e., the affine line $\mathbf{A}^1$ is the geometric quotient of $\mathbf{A}^4 - (\mathbf{A}^4)_{\text{reg}}$ by $G\ell(2, \mathbf{k})$. Similarly, for $n = 3$, one may classify orbits $\mathcal{O}(y)$, for $y \in (\mathbf{A}^9)_{\text{reg}}$, by polynomial equalities and inequalities, obtaining $\mathbf{A}^3$ as a geometric quotient. Inductively, a geometric quotient of X_{reg}—the set of regular points for the action of $G\ell(3, \mathbf{k})$ on the affine variety $X = \mathbf{A}^9 - (\mathbf{A}^9)_{\text{reg}}$—by $G\ell(3, \mathbf{k})$ is given by the union of a cubic and a surface in $\mathbf{A}^3$. Finally, a geometric quotient of $X - X_{\text{reg}}$ is given by $\mathbf{A}^1$.

Remarks. Although this description is fairly explicit, the problem of compactifying these quotients remains open. We have opted here to derive the polynomial inequalities involved (see Sect. 3 for a proof that such equalities and inequalities exist in general).

Using the dimension formula (2.1), one may obtain appropriate polynomial inequalities defining $(\mathbf{A}^{n^2})_{\text{reg}}$. It follows from an analysis of stability that one such inequality is not enough; i.e., $(\mathbf{A}^{n^2})_{\text{reg}}$ is not a principal affine open. However, it is easy to check

PROPOSITION 2.7. *$(\mathbf{A}^{n^2})_{\text{reg}} \xrightarrow{\phi} \mathbf{A}^n$ is a geometric quotient.*

DEFINITION 2.8. (i) $x \in \mathbf{A}^{n^2}$ is prestable if there is an invariant principal affine open containing x in which all orbits are closed.

(ii) $x \in \mathbf{A}^{n^2}$ is semistable provided there exists an invariant f with $x \in \mathbf{A}_f^{n^2}$.

(iii) $x \in \mathbf{A}^{n^2}$ is stable if there is an invariant f, with $x \in \mathbf{A}_f^{n^2}$ and for which all orbits in $\mathbf{A}_f^{n^2}$ are closed.

Remark. The properties of semistability and stability are taken here relative to the natural action of $G\ell(n, \mathbf{k})$ on (or the $G\ell(n, \mathbf{k})$-linearization of) the $\mathbf{k}$-algebra $\mathbf{k}[x_{ij}]$(of global sections of $o_{\mathbf{A}^{n^2}}$) (see [23, pp. 36–37]).

The corresponding point sets $(\mathbf{A}^{n^2})_{\text{pre}}$, $(\mathbf{A}^{n^2})_{\text{ss}}$, $(\mathbf{A}^{n^2})_{\text{s}}$ are open, and evidently $(\mathbf{A}^{n^2})_{\text{s}} \subset (\mathbf{A}^{n^2})_{\text{ss}} \cap (\mathbf{A}^{n^2})_{\text{pre}}$. By [23, Proposition 1.9], a geometric quotient, (Y, ψ), of $(\mathbf{A}^{n^2})_{\text{pre}}$ by $G\ell(n, \mathbf{k})$ exists, with ψ an affine map (see [23, p. 37]). Also, this proposition asserts that for any invariant open $U \subset \mathbf{A}^{n^2}$ possessing a geometric quotient, with the quotient map affine, $U \subset (\mathbf{A}^{n^2})_{\text{pre}}$. Now Mumford also shows [23, Theorem 1.10] that a categorical quotient (Z, μ) exists for $(\mathbf{A}^{n^2})_{\text{ss}}$ such that $(\mathbf{A}^{n^2})_{\text{s}} \xrightarrow{\mu} \tilde{Z}$ is a geometric quotient, for some open $\tilde{Z} \subset Z$, with Z quasi-projective. The situation here may be depicted as follows:

$$\begin{array}{ccccc}
(\mathbf{A}^{n^2})_{\text{s}} & \longrightarrow & (\mathbf{A}^{n^2})_{\text{pre}} & \longrightarrow & (\mathbf{A}^{n^2})_{\text{reg}} \\
\big\downarrow {\scriptstyle \phi|_{(\mathbf{A}^{n^2})_{\text{s}}}} & & \big\downarrow {\scriptstyle \phi|_{(\mathbf{A}^{n^2})_{\text{pre}}}} & & \big\downarrow {\scriptstyle \phi|_{(\mathbf{A}^{n^2})_{\text{reg}}}} \\
V & \longrightarrow & W & \longrightarrow & \mathbf{A}^n
\end{array} \tag{2.3}$$

where W, $V(\simeq\tilde{Z})$ are open in $\mathbf{A}^n$, $V \subsetneq W \subsetneq A^n$, and each vertical map is a geometric quotient. The inclusion $(\mathbf{A}^{n^2})_{\text{pre}} \subset (\mathbf{A}^{n^2})_{\text{reg}}$ follows from the closed orbit lemma and the fact that $(\mathbf{A}^{n^2})_{\text{pre}}$ is an invariant open set. Notice that, in view of this inclusion, $(\mathbf{A}^{n^2})_{\text{reg}}$ is a principal affine open if and only if $(\mathbf{A}^{n^2})_{\text{reg}} = (\mathbf{A}^{n^2})_{\text{pre}}$.

PROPOSITION 2.9. *The set of null forms,* $\mathbf{A}^{n^2} - (\mathbf{A}^{n^2})_{\text{ss}}$, *is precisely the set of nilpotent matrices. Thus,* $(\mathbf{A}^{n^2})_{\text{ss}}$ *has a categorical quotient which is not geometric.* $(\mathbf{A}^{n^2})_{\text{s}}$ *is the set of matrices with distinct eigenvalues, so that* $(\mathbf{A}^{n^2})_{\text{s}}$ *is the principal affine open defined by the invariant* $\delta(\phi, \phi')$, *the discriminant of the characteristic polynomial.*

Proof. Since the $\phi_\ell(x_{ij})$ generate $\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$, the null forms are just those matrices on which all the ϕ_ℓ vanish; i.e., those matrices with characteristic polynomial $\phi(x, \lambda) = \lambda^n$. As there exist nonclosed orbits $\mathcal{O}(x)$, for x not nilpotent, $(\mathbf{A}^{n^2})_{\text{ss}}$ cannot have a geometric quotient. Now

$$\delta(\phi, \phi') = \prod_{i<j} (\lambda_i - \lambda_j)^2,$$

where $\{\lambda_1, ..., \lambda_n\}$ is the spectrum of x. In particular, $\delta(\phi, \phi')$ is a symmetric function in the eigenvalues of x and so $\delta(\phi, \phi') \in \mathbf{k}[\phi_1, ..., \phi_n]$, where the ϕ_ℓ are the elementary symmetric functions of the λ_i; i.e., the coefficients of $\phi(x, \lambda)$. Hence, $\delta(\phi, \phi') \in \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$. In particular, each matrix x with distinct eigenvalues satisfies (iii) of Definition 2.8, since $x \in \mathbf{A}_\delta^{n^2}$ and the orbits in $\mathbf{A}_\delta^{n^2}$ are closed (see the remark following Corollary 2.4). To see that $(\mathbf{A}^{n^2})_{\text{s}} \subset \mathbf{A}_\delta^{n^2}$, notice that $(\mathbf{A}^{n^2})_{\text{s}}$ may also be characterized as those $x \in (\mathbf{A}^{n^2})_{\text{reg}}$ for which $\mathcal{O}(x)$ is closed (see [24, p. 185, Remark]). Q.E.D.

The open set V of diagram (2.3) may be described explicitly and, in fact, we have already done so for the case $n = 2$. Notice that, in this case the open set W lies between the colocus of $f(x, y) = x^2 - 4y$ and the entire plane $\mathbf{A}^2$. If this containment is proper, then W must be $\mathbf{A}^{n^2}$ less a finite number of points, while one can argue directly that W must miss the origin. More generally, since W is quasi-projective, qualitative results [23, Converse 1.13, p. 4] imply that the prestable points are stable relative to some (fixed) action of $G\ell(n, \mathbf{k})$ on $\mathbf{k}[x_{ij}]$. Thus, one might hope for the inclusion $(\mathbf{A}^{n^2})_{\text{pre}} \subset (\mathbf{A}^{n^2})_{\text{s}}$.

PROPOSITION 2.10. *If* $f \in \mathbf{k}[x_{ij}]$ *and* $\mathbf{A}_f^{n^2}$ *is* $G\ell(n, \mathbf{k})$*-invariant, then* $\mathbf{A}_f^{n^2} = \mathbf{A}_{f_1}^{n^2}$, *with* $f_1 \in \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$.

COROLLARY 2.11. $(\mathbf{A}^{n^2})_{\text{s}} = (\mathbf{A}^{n^2})_{\text{pre}}$.

Proof. Proposition 2.10 renders (i) and (iii) of Definition 2.8 equivalent.

COROLLARY 2.12. $(\mathbf{A}^{n^2})_{\text{reg}}$ *is not a principal affine open.*

Proof. If this were the case, $(\mathbf{A}^{n^2})_{\text{reg}} \to^{\Phi} \mathbf{A}^n$ would be an affine map. Thus $(\mathbf{A}^{n^2})_{\text{reg}} = (\mathbf{A}^{n^2})_{\text{pre}}$, by our previous remarks. The containments obtained so far yield

$$(\mathbf{A}^{n^2})_{\text{pre}} = (\mathbf{A}^{n^2})_{\text{s}} \subset (\mathbf{A}^{n^2})_{\text{ss}} \cap (\mathbf{A}^{n^2})_{\text{reg}} \subset (\mathbf{A}^{n^2})_{\text{reg}}.$$

However, the right-hand containment is proper, since there exists a nonderogatory, nilpotent matrix. Q.E.D.

Proof of Proposition 2.10. Let $f = p_1^{e_1} \cdots p_r^{e_r}$ be the prime factorization of f and consider $f_1 = p_1 \cdots p_r$. Clearly $\mathbf{A}_f^{n^2} = \mathbf{A}_{f_1}^{n^2}$, and the same is true for the loci, $Z_f = Z_{f_1}$. Now $Z_{f_1} = \bigcup_{i=1}^{r} Z_{p_i}$ is a decomposition of Z_{f_1} into an irredundant union of irreducible varieties. As is well known (see [9, p. 20]), such a decomposition is unique (up to permutation). Since $\mathbf{A}_{f_1}^{n^2}$ is $G\ell(n, \mathbf{k})$-invariant, Z_{f_1} is invariant as well. Thus if $g \in G\ell(n, \mathbf{k})$, then $g(Z_{p_i}) g^{-1} = Z_{p_j}$, for each $i = 1, \ldots, r$ and for some j, by the uniqueness of the decomposition. Alternately said, $Z_{(p_i^g)} = Z_{p_j}$, and by the weak Nullstellansatz (see [2, p. 47]) one obtains the relations $(p_i^g) \mid p_g^{t_i}$ and $p_j \mid (p_i^g)^{s_i}$ for positive integers s_i, t_i. Since $G\ell(n, \mathbf{k})$ acts on $\mathbf{k}[x_{ij}]$ automorphically, each p_i^g is irreducible as well, so that one actually has $p_i^g \mid p_j$ and $p_j \mid (p_i^g)$. Consequently, $p_i^g = \alpha_{ij}(g) p_j$ where $\alpha_{ij}(g)$ is a unit in $\mathbf{k}[x_{ij}]$, i.e., a nonzero scalar, depending on g. In fact,

$$f_1^g = \chi(g) f_1, \qquad \text{for} \quad \chi(g) \in \mathbf{k} - \{0\}, \tag{2.4}$$

and one may verify that $\chi(g_1 g_2) = \chi(g_1) \cdot \chi(g_2)$. In other words, χ is a character of $G\ell(n, \mathbf{k})$ and f_1 is a semi-invariant, or f_1 is a relative invariant of weight χ. Furthermore, Eq. (2.4) shows that χ is a rational character of $G\ell(n, \mathbf{k})$ and it is known that all such characters are expressible as $\det(g)^t$, for some integer t (see [8, p. 22]). If $g = \lambda \cdot I$, then clearly

$$f_1(x) = f_1{}^g(x) = \lambda^{nt} f_1(x).$$

Choosing $x \in \mathbf{A}_{f_1}^{n^2}$ and λ not to be a root of unity yields $t = 0$; i.e., χ is the trivial character. Hence $\mathbf{A}_f^{n^2} = \mathbf{A}_{f_1}^{n^2}$, with $f_1 \in \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$. Q.E.D.

Remarks. 1. In the light of Proposition 2.9, we can view the problem as reduced into two parts: an analysis of the quasi-projective (rational) data defining $(\mathbf{A}^{n^2})_{\mathrm{SS}}$ and that of finding criteria for separating null forms. Actually, this latter, "exceptional" case is the most easily solved. For in the nilpotent (or even the one eigenvalue) case, Weyr's invariants are algebraically computable. Remarkably, a refined version of this criteria is crucial in the general case. In fact at one stage in the proof of the Main Theorem, we reduce to the nilpotent case and use this criteria. This explains the presence of the higher powers which appear in condition (ii) of the Main Theorem.

2. The motivation for condition (ii), with $k = 1$, comes from the first part of this reduction. Inductively, as a "first approximation" to our Main Theorem, consider $(\mathbf{A}^{n^2})_{\mathrm{reg}} \cap (\mathbf{A}^{n^2})_{\mathrm{SS}}$ (or $(\mathbf{A}^{n^2})_{\mathrm{reg}}$, filling in the origin) rather than $(\mathbf{A}^{n^2})_{\mathrm{S}}$ or, what is the same here, $(\mathbf{A}^{n^2})_{\mathrm{pre}}$. Now in the light of Corollary 2.12, it is much more efficient (both theoretically and computationally) to define $(\mathbf{A}^{n^2})_{\mathrm{reg}}$ by the single algebraic condition: rank $\mathrm{ad}(x) = n^2 - n$. However (for precisely the reasons given in the case $n = 4$), the invariant rank $\mathrm{ad}(x)$ is not sensitive enough and, using a two-variable analog of $\mathrm{ad}(x)$ intimately connected with a resultant calculation, we consider a related geometric problem—that of looking for the graph of the similarity relation in $\mathbf{A}^{n^2} \times \mathbf{A}^{n^2}$.

In the next section, it is shown that these observations, together with the key observation (see [11]) that, by representing $\mathrm{ad}(x)$ on $(\mathbf{k}^n \otimes_{\mathbf{k}} \mathbf{k}^n)$, a great deal of the block structure of x may be recovered using the Clebsch–Gordon formula for representations of $s\ell_2$, may be formalized so as to obtain effective criteria for deciding the equality $\mathcal{O}(x) = \mathcal{O}(y)$. Included in Section 4 are applications of the dimension formula (2.1) and the above constructions to the global analysis of finite-dimensional, time-invariant linear dynamical systems.

3. The Constructibility Method

In Section 2 we saw the difficulties one encounters in trying to view the similarity problem as a moduli problem. Here we look at the problem from a decidability point of view and find an explicit solution (Theorem 3.6) which is suggested by the discussion to follow.

In any decent algebraic classification problem one has a collection of objects

V and a notion of isomorphism, a group G, and an action of G on V for which orbits are isomorphism classes. Depending on one's philosophy, the classification problem can be viewed in several ways. From one point of view, one would like to have for each member of V a complete list of its isomorphism invariants (this is the study of the structure of V/G). From another point of view, one could consider the classification problem to be solved, if given v, w in V, one could effectively decide whether or not they are isomorphic; that is, one wants a suitable description of the subset $\mathscr{I}$ of $V \times V$ consisting of all pairs (v, w) for which $v \simeq w$, the graph of the relation. By a theorem of Chevalley, such a result is available in many situations.

PROPOSITION 3.1. *Let G be an algebraic group acting on the variety V by a morphism $\alpha : G \times V \to V$. Then $\mathscr{I} = \{(v, w) | \ v \equiv_G w\}$ is a constructible subset of $V \times V$.*

Proof. $\mathscr{I}$ is the image of the morphism $G \times V \to V \times V$ given by $(g, v) \to (\alpha(g, v), v)$. Chevalley's theorem [22, p. 97] asserts that the image of a morphism is constructible.

For the remainder of this section we are concerned with producing such a decidability criterion (i.e., algebraic description of $\mathscr{I}$) for the similarity problem. We retain the notation of Section two for Jordan block matrices and diagonal sums. In addition, if A is an $n \times n$ matrix and B is $m \times m$, instead of $A \otimes I_m - I_n \otimes B$, we simply write $A \otimes 1 - 1 \otimes B$. Note that if A is in End(V) for some finite-dimensional vector space V, there is a natural isomorphism $V \otimes V^* \simeq$ End(V), under which ad(A) (acting on End(V)) is identified with $A \otimes 1 - 1 \otimes {}^tA$ (t denoting transpose). Since A is similar to its transpose, $A \otimes 1 - 1 \otimes A$ is just a noncanonical version of ad(A).

Now the rational data (polynomial equalities and inequalities) achieved in dimensions 2 and 3 in Section 2, is related to the rational data we seek here that is assured by Chevalley's theorem. One should expect that the equalities are pretty much in hand. In fact, it is not hard to show, using [22, Corollary 1, p. 114], that the Zariski-closure of $\{(A, B) | \ A \sim B\}$ in $\mathbf{k}^{n^2} \times \mathbf{k}^{n^2}$ is the set of common zeros of the functions $\phi_i(x) - \phi_i(y)$, $i = 1, \ldots, n$. It remains to find the inequalities. The most obvious way inequalities arise in matrix theory is through rank computations. We first explore this idea by considering the null forms. The similarity problem is completely solved by a result of Weyr, valuable in its own right.

LEMMA 3.2. rank$(J(0; n))^k = n - k$ *for* $k \leqslant n$, *and is zero for* $k \geqslant n$.

PROPOSITION 3.3 (Weyr). *Let N_1 and N_2 be nilpotent $n \times n$ matrices. Then N_1 is similar to N_2 if and only if* rank $N_1^k =$ rank N_2^k *for all* $k \leqslant n$.

Proof. The necessity of the conditions is clear. For the sufficiency write

$\oplus_{i=1}^{r} J(0; n_i)$, $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_r$, and $\oplus_{i=1}^{s} J(0; m_i)$, $m_1 \leqslant m_2 \leqslant \cdots \leqslant m_s$, for the Jordan forms of N_1 and N_2, respectively. If $n_r \neq m_s$ we can assume $n_r > m_s$ without harm. Then $N_1^{n_r-1} \neq 0$ while $N_2^{n_r-1} = 0$. This contradiction to the rank assumptions forces $m_s = n_r$, and one can proceed in an obvious way to finish the proof by induction on n.

The next two results are technical and needed in the proof of our main theorem.

LEMMA 3.4. *If A and B are similar matrices, so are $A \otimes 1 - 1 \otimes A$, $B \otimes 1 - 1 \otimes B$, and $A \otimes 1 - 1 \otimes B$. Also, if λ_1 and λ_2 are field elements and m and n are natural numbers, one has that*

$$J(\lambda_1; n) \otimes 1 - 1 \otimes J(\lambda_2; m) \sim \bigoplus_{t=1}^{\min(m,n)} J(\lambda_1 - \lambda_2; m + n + 1 - 2t). \qquad (3.1)$$

Proof. The first result is very easy, and the second is due to Roth [26]. Gauger has shown in [11] that this is really just the Clebsch–Gordon Formula for $s\ell_2$. This is where our criterion may fail in finite characteristics. The representation theory of $s\ell_2$ over these fields is known to be quite different. For example, over $\mathbf{Z}_2$ one may check that $J(1; 2) \otimes 1 - 1 \otimes J(1; 2)$ is similar to $J(0; 2) \oplus J(0; 2)$, not $J(0; 3) \oplus J(0; 1)$ as Roth's result would predict.

LEMMA 3.5. (Reduction lemma). *Suppose A and B are both $n \times n$ matrices, suppose $A = A' \oplus D$ and $B = B' \oplus D$, and suppose*

$$\begin{aligned} \operatorname{rank}(A \otimes 1 - 1 \otimes A)^k + \operatorname{rank}(B \otimes 1 - 1 \otimes B)^k \\ - 2 \operatorname{rank}(A \otimes 1 - 1 \otimes B)^k = 0. \end{aligned} \qquad (3.2)$$

Then one has a similar identity replacing A by A' and B by B'.

Proof. Expanding sums across tensors one gets

$$\begin{aligned} &(A \otimes 1 - 1 \otimes A)^k \\ &\quad = (A' \otimes 1 - 1 \otimes A')^k \oplus (D \otimes 1 - 1 \otimes A')^k \oplus (A' \otimes 1 - 1 \otimes D)^k \\ &\quad\quad \oplus (D \otimes 1 - 1 \otimes D)^k, \end{aligned}$$

$$\begin{aligned} &(B \otimes 1 - 1 \otimes B)^k \\ &\quad = (B' \otimes 1 - 1 \otimes B')^k \oplus (D \oplus 1 - 1 \otimes B')^k \oplus (B' \otimes 1 - 1 \otimes D)^k \\ &\quad\quad \oplus (D \otimes 1 - 1 \otimes D)^k, \end{aligned}$$

$$\begin{aligned} &(A \otimes 1 - 1 \otimes B)^k \\ &\quad = (A' \otimes 1 - 1 \otimes B')^k \oplus (D \otimes 1 - 1 \otimes B')^k \oplus (A' \otimes 1 - 1 \otimes D)^k \\ &\quad\quad \oplus (D \otimes 1 - 1 \otimes D)^k, \end{aligned}$$

first for $k = 1$, and then easily for all k. Notice that $\operatorname{rank}(D \otimes 1 - 1 \otimes A')^k = \operatorname{rank}(A' \otimes 1 - 1 \otimes D)^k$ and similarly for B', since $A' \otimes 1 - 1 \otimes D$ is similar to the negative of $D \otimes 1 - 1 \otimes A'$ (see Eq. (3.1)). One obtains the desired result by taking ranks on both sides of the three equations (rank distributes across $\oplus$), and then subtracting two times the last equation from the sum of the first two. The left side is zero due to (3.2).

Finally we are in a position to give our main result.

THEOREM 3.6. *Let A and B be $n \times n$ matrices over a field of characteristic zero. Then A is similar to B if and only if*

(i) *A and B have the same characteristic polynomial, and*

(ii) *for all even $k \leqslant 2n - 2$,*

$$\operatorname{rank}(A \otimes 1 - 1 \otimes A)^k = \operatorname{rank}(B \otimes 1 - 1 \otimes B)^k = \operatorname{rank}(A \otimes 1 - 1 \otimes B)^k.$$

Remarks 1. From the dimension formula (2.1) it is clear that the dimension of a similarity class as a variety is completely determined by the sizes of the Jordan blocks. As we pointed out there, this dimension is the same as the rank of $\operatorname{ad}(A)$. We have also pointed out that $A \otimes 1 - 1 \otimes A$ is just a noncanonical version of $\operatorname{ad}(A)$. Thus, the motivation for (ii) lies in orbital dimension calculations and the solution for null forms (Weyr's theorem).

2. In the proof of the theorem we actually replace the second condition by the weaker condition (3.2) since it is carried through the induction step of our argument more easily (see the reduction lemma).

Proof. The necessity of the conditions is obvious. As for the sufficiency, (i) says that the eigenvalues and their multiplicities are identical in A and B. We can assume A and B are in Jordan form since (i) and (3.2) are unaffected in replacing A and B by something similar. We will see that (3.2) (hence (ii)) guarantees that the block sizes for each eigenvalue in A and B are identical.

So let $\lambda_1, \ldots, \lambda_r$ be a pairwise distinct list of the eigenvalues, and let $m_1, \ldots, m_r$ be their multiplicities; $\sum_i m_i = n$. We can write $A = \bigoplus_{i=1}^r (\bigoplus_{j=1}^{a_i} J(\lambda_i\,;\, p_{ij}))$, $B = \bigoplus_{i=1}^r (\bigoplus_{k=1}^{b_i} J(\lambda_i\,;\, q_{ik}))$ where $p_{i1} \leqslant p_{i2} \leqslant \cdots \leqslant p_{ia_i}$ and $q_{i1} \leqslant q_{i2} \leqslant \cdots \leqslant q_{ib_i}$ for all i.

Keeping in mind (3.1), replace A and B by their sums in terms of Jordan blocks and expand $A \otimes 1 - 1 \otimes A$, $B \otimes 1 - 1 \otimes B$, and $A \otimes 1 - 1 \otimes B$. One sees that each matrix is of the form $P \oplus N$ where P is nonsingular and of size $(n^2 - \sum_{i=1}^r m_i^2) \times (n^2 - \sum_{i=1}^r m_i^2)$, and N is nilpotent and of size $(\sum_{i=1}^r m_i^2) \times (\sum_{i=1}^r m_i^2)$. Now the rank of $(P \oplus N)^k = P^k \oplus N^k$ is just a constant, $n^2 - \sum_{i=1}^r m_i^2$, plus the rank of N^k.

Write $A \otimes 1 - 1 \otimes A = P_1 \oplus N_1$, $B \otimes 1 - 1 \otimes B = P_2 \oplus N_2$ and

$A \otimes 1 - 1 \otimes B = P_3 \oplus N_3$ where the P_i and N_i are as above. By (3.2) and the remarks of the preceding paragraph one obtains

$$\operatorname{rank} N_1{}^k + \operatorname{rank} N_2{}^k - 2 \operatorname{rank} N_3{}^k = 0 \tag{3.3}$$

for all even $k \leqslant 2n - 2$.

Now the N_i can be described more explicitly;

$$\begin{aligned}
N_1 &= \bigoplus_{i=1}^{r} \left(\bigoplus_{k,\ell=1}^{a_i} J(\lambda_i; p_{ik}) \otimes 1 - 1 \otimes J(\lambda_i; p_{i\ell}) \right) \\
&= \bigoplus_{i=1}^{r} \bigoplus_{k,\ell=1}^{a_i} \bigoplus_{t=1}^{\min(p_{ik}; p_{i\ell})} J(0; p_{ik} + p_{i\ell} + 1 - 2t), \\
N_2 &= \bigoplus_{i=1}^{r} \left(\bigoplus_{k,\ell=1}^{b_i} J(\lambda_i; q_{ik}) \otimes 1 - 1 \otimes J(\lambda_i; q_{i\ell}) \right) \\
&= \bigoplus_{i=1}^{r} \bigoplus_{k,\ell=1}^{b_i} \bigoplus_{t=1}^{\min(q_{ik}, q_{i\ell})} J(0; q_{ik} + q_{i\ell} + 1 - 2t), \\
N_3 &= \bigoplus_{i=1}^{r} \left(\bigoplus_{k=1\ \ell=1}^{a_i b_i} J(\lambda_i; p_{ik}) \otimes 1 - 1 \otimes J(\lambda_i; q_{i\ell}) \right) \\
&= \bigoplus_{i=1}^{r} \bigoplus_{k=1\ \ell=1}^{a_i b_i} \bigoplus_{t=1}^{\min(p_{ik}, q_{i\ell})} J(0; p_{ik} + q_{i\ell} + 1 + 2t).
\end{aligned} \tag{3.4}$$

Let ℓ_1 be the largest block size in A and let ℓ_2 be the largest block size in B. Without harm we can suppose $\ell_1 \geqslant \ell_2$. Suppose

there is no eigenvalue λ_i having the same largest block size in both A and B. (3.5)

Then the largest block size in N_1 is $2\ell_1 - 1$, the largest in N_2 is $2\ell_2 - 1$, while (3.4) and (3.5) imply that the largest block size in N_3 is $\leqslant 2\ell_1 - 2$. Then for $k = 2\ell_1 - 2$, $N_3{}^k = 0$ but $N_1{}^k \neq 0$. This contradicts (3.3). Hence (3.5) is false. So for some i the largest λ_i blocks in A and B are identical. Apply the reduction lemma with D equal to this block and proceed by induction on n to finish the argument. Q.E.D.

Remark. It is interesting to note that the techniques developed here hint at definitions of resultants and discriminants which are seemingly easier to work with than the classical ones.

If $p(x)$ is a monic polynomial of degree n, we let $C(p(x))$ denote the corresponding $n \times n$ companion matrix. For monic polynomials $p(x)$ and $q(x)$ it is easily checked that (i) $\det(C(p(x)) \otimes 1 - 1 \otimes C(q(x))$ vanishes if and only if $p(x)$ and $q(x)$ have a common root (the eigenvalues of $C(p(x)) \otimes 1 - 1 \otimes C(q(x))$ are the differences of roots of $p(x)$ with roots of $q(x)$), and

(ii) that this determinant is in fact equal to the resultant of $p(x)$ and $q(x)$. Setting $q(x) = p'(x)$ in the determinant thus gives the discriminant of $p(x)$. In fact, one can show

$$\begin{aligned} &\operatorname{rank}(C(p(x)) \otimes 1 - 1 \otimes C(q(x))) \\ &\quad = \deg(p(x)\, q(x)) - \deg \text{g.c.d.}\, \{p(x), q(x)\} \\ &\quad = \deg \text{l.c.m.}\{p(x), q(x)\}. \end{aligned} \tag{3.6}$$

Proof of 3.6. Let $p(x) = \prod_{i=1}^{r} (x - \lambda_i)^{e_i}$, $q(x) = \prod_{j=1}^{s} (x - \omega_j)^{f_j}$ be the prime factorizations of $p(x)$ and $q(x)$. Then $C(p(x)) \sim \bigoplus_{i=1}^{r} J(\lambda_i ; e_i)$ and $C(q(x)) \sim \bigoplus_{j=1}^{s} J(\omega_j ; f_j)$. The rank in question is $\deg(p(x)\, q(x))$ minus the number of 0-blocks which occur in the Jordan form. Now 0-blocks can occur for each i and j such that $\lambda_i = \omega_j$. In this case

$$J(\lambda_i; e_i) \otimes 1 - 1 \otimes J(\omega_j; f_j) \sim \bigoplus_{t=1}^{\min(e_i, f_j)} J(0; e_i + f_j + 1 - 2t)$$

contributes $\min(e_i, f_j)$ 0-blocks; $\min(e_i, f_j)$ is the power to which $(x - \lambda_i) = (x - \omega_j)$ occurs in the greatest common divisor.

4. Moduli of Certain Linear Dynamical Systems

The classification of linear control systems, defined over an arbitrary field $\mathbf{k}$, has been discussed by authors in both the engineering and mathematical literature (see especially [4, 5, 12, 16, 17]). In this section we construct a moduli space for a class of such systems in quite a different fashion. We also indicate some of the elementary applications one can derive from a global analysis of these systems.

Roughly speaking, a linear dynamical system is dependent on three parameters—input, state, and output vectors. More precisely, the evolution (i.e., the change of state and output) of a system is assumed to be linear in these parameters and, thus, the only possibilities depend on

(i) three vector spaces U, X, Y (the input, state, and output spaces), with $\dim_{\mathbf{k}}(U) = m$, $\dim_{\mathbf{k}}(X) = n$, $\dim_{\mathbf{k}}(Y) = p$, and

(ii) four maps, $F \in \operatorname{Hom}_{\mathbf{k}}(X, X)$, $G \in \operatorname{Hom}_{\mathbf{k}}(U, X)$, $H \in \operatorname{Hom}_{\mathbf{k}}(X, Y)$, and $K \in \operatorname{Hom}_{\mathbf{k}}(U, Y)$.

The equations of evolution, in matrix form, for a discrete time system over

$$x(t+1) = Fx(t) + Gu(t), \tag{4.1}$$

$$y(t) = Hx(t) + Ku(t). \tag{4.2}$$

In essence, we may take the quadruple (F, G, H, K), with (4.1) and (4.2) implicitly understood, as our definition of a finite-dimensional, time-invariant, discrete-time, linear dynamical system—referred to more succintly as an LDS. Viewing an LDS in terms of the usual "blackbox" picture, one observes that, although by numbering the input and output terminals it is possible to choose bases for U and Y naturally, one cannot choose a canonical basis for the state space X. Thus, a natural notion of equivalence for such systems is induced by a change of basis in X.

DEFINITION 4.1. $(F, G, H, K) \sim (F_1, G_1, H_1, K_1)$ if and only if $(F_1, G_1, H_1, K_1) = (gFg^{-1}, gG, Hg^{-1}, K)$, for some $g \in G\ell(n, \mathbf{k})$.

Remarks. 1. The orbits of this action are precisely the classes partitioned by system-theoretic equivalence; i.e., each orbit is the class of systems with the same "input–output" pairs (see [18]), for controllable and observable systems.

2. Obviously, we may set $K = 0$ without altering the notion of equivalence.

Now, in the classification of LDS's, it has been customary to restrict attention to a large class of particularly nice systems. These are the important systems for which any state vector x_1 may be reached from a fixed but arbitrary state vector x_0 by a suitable choice of inputs. It is well known that this property is equivalent to the following:

DEFINITION 4.2. (F, G, H, K) is completely controllable provided the matrix $(G \mid FG \mid \cdots \mid F^{n-1}G)$ has rank n.

Remark. Notice that rank $(G \mid FG \mid \cdots \mid F^{n-1}G) = \text{rank } (G \mid FG \mid \cdots \mid F^{j}G)$ for all $j \geqslant n-1$, by the Cayley–Hamilton theorem.

While we are about it, set

$$\tilde{\Sigma} = \{(F, G, H, 0) : (F, G, H, 0) \text{ is completely controllable}\},$$
$$\tilde{\Sigma}_0 = \{(F, G, 0, 0) \in \tilde{\Sigma}\}, \Sigma = \tilde{\Sigma}/G\ell(n, \mathbf{k}), \text{ and } \Sigma_0 = \tilde{\Sigma}_0/G\ell(n, \mathbf{k}).$$

The quotient $\tilde{\Sigma} \to \Sigma$ has been constructed by several authors in the scalar input–output case, i.e., when $m = p = 1$. Actually, Hermann [12] considered a slightly different class of systems (see Corollary 4.14). Kalman [16, 17] sketched a construction for Σ in the general case, but his techniques have developed some trouble. Byrnes and Hurt have constructed this quotient [4, 5] and have shown $\tilde{\Sigma} \to \Sigma$ to be geometric, obtaining results differing from those announced in [16, 17] when $m > 1$. Here we construct the quotient for $m = 1$, p arbitrary, based on the results in Section 2 and in [24]. We first consider $\tilde{\Sigma}_0$ in terms of the action of $G\ell(n, \mathbf{k})$ on $\mathbf{A}^{n^2+n}$. By Mumford's theorem [23], there is a categorical quotient for this action and, as in Section 2, we identify this quotient by finding the ring of invariants. Since the coordinate ring of $\mathbf{A}^{n^2+n}$ is $\mathbf{k}[x_{ij}] \otimes_{\mathbf{k}} \mathbf{k}[x_1, ..., x_n]$, the next lemma shows that we have a surprising isomorphism:

$$\begin{aligned}(\mathbf{k}[x_{ij}] \otimes_{\mathbf{k}} \mathbf{k}[x_1, ..., x_n])^{G\ell(n,\mathbf{k})} &\simeq \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})} \otimes_{\mathbf{k}} \mathbf{k}[x_1, ..., x_n]^{G\ell(n,\mathbf{k})} \\ &\simeq \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}.\end{aligned}$$

LEMMA 4.3. *If h: $\mathbf{A}^{n^2+n} \to \mathbf{A}^1$ is invariant, then h depends only on F.*

Proof. Since $G\ell(n, \mathbf{k})$ preserves degrees, we can assume that $h = \sum_i p_i(F)\, h_i(G)$ where each $h_i(G)$ is a monomial of degree d. Setting $g = \alpha \cdot Id$, where $\alpha^d \neq 1$, one has

$$h = h^g = \sum_i p_i^g(F)\, h_i^g(G) = \sum_i p_i(F)\, \alpha^d h_i(G) = \alpha^d h.$$

Thus $d = 0$.

COROLLARY 4.4. *The map $\phi_1 : \mathbf{A}^{n^2+n} \to \mathbf{A}^n$, where $\phi_1(F, G) = \phi(F)$ as in Section 2, is a categorical quotient for the action of systems—theoretic equivalence.*

We can use Lemma 4.3 in a slightly different manner. The map ϕ_1 factors as $\phi \circ \mathrm{proj}_1$, where $\mathrm{proj}_1 : \mathbf{A}^{n^2+n} \to \mathbf{A}^{n^2}$ is defined by $\mathrm{proj}_1(F, G) = F$. In order to show that $\phi_1 : \mathbf{A}^{n^2+n} \to \mathbf{A}^n$ is a categorical quotient, in light of Proposition 2.6 it is sufficient to show that any invariant h: $\mathbf{A}^{n^2+n} \to \mathbf{A}^1$ descends to a unique map $h_1 : \mathbf{A}^{n^2} \to \mathbf{A}^1$ satisfying $h_1 \circ \mathrm{proj}_1 = h$. Uniqueness follows since h_1 is just the restriction of h to the subvariety $\mathbf{A}^{n^2} \times \{0\}$ of $\mathbf{A}^{n^2+n}$, while the condition $h_1 \circ \mathrm{proj}_1 = h$ is just Lemma 4.3. We can use this observation for the subset $\tilde{\Sigma}_0 \subset \mathbf{A}^{n^2+n}$. First claim that $\phi_1 : \tilde{\Sigma}_0 \to \mathbf{A}^n$ is a quotient. To see this, consider the diagram

$$\begin{tikzcd} \Sigma_0 \arrow[r, "h"] \arrow[d, "\mathrm{proj}_1"'] & \mathbf{A}^1 \\ (\mathbf{A}^{n^2})_{reg} \arrow[ur, "h_1"] \arrow[d, "\phi"'] & \\ \mathbf{A}^n \arrow[uur, "h_2"'] & \end{tikzcd} \tag{4.3}$$

for h invariant, as above. It is easy to see that h descends to a map $h_1 : (\mathbf{A}^{n^2})_{\text{reg}} \to \mathbf{A}^1$ and, by Lemma 4.3, $h_1 \circ \text{proj}_1 = h$. Moreover, by Proposition 2.7, $(\mathbf{A}^{n^2})_{\text{reg}} \to^{\phi} \mathbf{A}^n$ is a geometric quotient, so h_1 descends to a unique map h_2, satisfying $h_2 \circ \phi = h_1$. As before, composing this last identity with proj_1 shows that $\tilde{\Sigma}_0 \to^{\phi_1} \mathbf{A}^n$ is a categorical quotient. Now one may derive, from Proposition 2.7, that this quotient is geometric. This is also a consequence of standard results on free actions and follows, as well, from an analysis of stability. In order to get a handle on the set of points which are regular for this action, we derive a dimension formula for orbits. This formula is also essential in our classification of LDS's in Section 5.

PROPOSITION 4.5. $\dim \mathcal{O}(F, G) = \dim \mathcal{O}(F) + \dim(Z(F))(G))$.

Proof. As in Proposition 2.3, we need only compute the derivative (at the identity) of the map

$$\alpha_{(F,G)} : G\ell(n, \mathbf{k}) \to \mathcal{O}(F, G),$$

defined by

$$\alpha_{(F,G)}(g) = (gFg^{-1}, gG).$$

Here one has, for L in the Lie algebra of $G\ell(n, \mathbf{k})$, viz, $M(n, \mathbf{k})$,

$$d\alpha_{(F,G)}(Id)(L) = ([F, L], LG).$$

Thus, $\operatorname{rank} d\alpha_{(F,G)}(Id) = \operatorname{rank} \operatorname{ad}(F) + \dim(Z(F)(G))$. Since $\operatorname{rank} \operatorname{ad}(F) = \dim \mathcal{O}(F)$, by Proposition 2.3, one obtains the equality

$$\dim \mathcal{O}(F, G) = \dim \mathcal{O}(F) + \dim(Z(F)(G)). \tag{4.4}$$

Q.E.D.

Remark. This enables us to considerably refine the (obvious) inequality $\dim \mathcal{O}(F, G) \leqslant n^2$. Explicitly, one can bound the right-hand side of (4.4) as follows: $\dim \mathcal{O}(F) \leqslant n^2 - n$ (see Proposition 2.5) and $\dim(Z(F)(G)) \leqslant n$, since $G \in \mathbf{k}^n$.

THEOREM 4.6. $(\mathbf{A}^{n^2+n})_{\text{reg}} = (\mathbf{A}^{n^2+n})_{\text{pre}} = \tilde{\Sigma}_0$.

Proof. We first show $\tilde{\Sigma}_0 \subset (\mathbf{A}^{n^2+n})_{\text{pre}}$. Now the function $f(F, G) = \det(G \mid FG \mid \cdots \mid F^{n-1}G)$ is clearly a relative invariant of weight $\det(\cdot)$, i.e., $f(gFg^{-1}, gG) = \det(g) \cdot f(F, G)$. In particular $\mathbf{A}_f^{n^2+n} = \tilde{\Sigma}_0$ is an invariant principal affine open, so we need only check that the action of $G\ell(n, \mathbf{k})$ on $\tilde{\Sigma}_0$ is closed. In fact, we claim that the action is free. Thus, assuming the claim, one has that all orbits in $\tilde{\Sigma}_0$ have the same dimension, viz n^2, so that all orbits are closed, by the closed orbit lemma. For a proof of the claim, consider the map $R\colon \tilde{\Sigma}_0 \to M_*(n, n+1)$, defined by $R(F, G) = (G \mid FG \mid \cdots \mid F^nG)$, where $M_*(n, n+1)$ denotes the set of $n \times (n+1)$ matrices of rank n. Notice that R is injective, since G is a cyclic vector for F. Moreover R is an intertwining operator for the action of $G\ell(n, \mathbf{k})$ on $\tilde{\Sigma}_0$ and the left multiplication of $G\ell(n, \mathbf{k})$ on $M_*(n, n+1)$; i.e., the diagram

$$\begin{tikzcd} \tilde{\Sigma}_0 \arrow[r, "R"] \arrow[d, "g"] & M_*(n, n+1) \arrow[d, "g"] \\ \tilde{\Sigma}_0 \arrow[r, "R"] & M_*(n, n+1) \end{tikzcd} \tag{4.5}$$

is commutative, for all $g \in G\ell(n, \mathbf{k})$. Thus, $(gFg^{-1}, gG) = (F, G)$ if and only if $gR(F, G) = R(F, G)$. However, since the $n \times n$ submatrix $(G \mid FG \mid \cdots \mid F^{n-1}G)$ is invertible, $gR(F, G) = R(F, G)$ implies $g = Id$. As in Section 2, $(\mathbf{A}^{n^2+n})_{\mathrm{pre}} \subset (\mathbf{A}^{n^2+n})_{\mathrm{reg}}$, so that it is enough to show $(\mathbf{A}^{n^2+n})_{\mathrm{reg}} \subset \tilde{\Sigma}_0$. Now if $(F, G) \in (\mathbf{A}^{n^2+n})_{\mathrm{reg}}$, then $\dim \mathcal{O}(F, G) = n^2$, since this is true for each $(F, G) \in \tilde{\Sigma}_0 \subset (\mathbf{A}^{n^2+n})_{\mathrm{reg}}$. This restriction on $\dim \mathcal{O}(F, G)$ gives the values $\dim \mathcal{O}(F) = n^2 - n$, and $\dim(Z(F)(G)) = n$ for the right-hand side of (4.4). By Proposition 2.5, the first equality implies that F is nonderogatory and, in particular, we may replace $Z(F)$ by $\mathbf{k}[F]$. The second equality now states, by a dimension argument, that $\mathbf{k}[F](G) = \mathbf{k}^n$, from which it follows, easily, that G is cyclic vector for F; i.e., $(F, G) \in \tilde{\Sigma}_0$. Q.E.D.

Corollary 4.7. *$\tilde{\Sigma}_0 \xrightarrow{\phi_1} \Sigma_0 \simeq \mathbf{A}^n$ is a (universal) geometric quotient for the free action of $G\ell(n, \mathbf{k})$ in $\tilde{\Sigma}_0$, such that ϕ_1 is affine.*

Remark. Notice that the intertwining operator may also be used to exhibit Σ_0 as $\mathbf{A}^n$. Explicitly, since R intertwines the actions involved, one may show there exists an imbedding of Σ_0 into $G\ell(n, \mathbf{k})/M_*(n, n+1)$. This latter quotient is well known to exist and, in fact, is the Grassmann variety of n-planes in $(n+1)$-space, denoted $\mathbf{Gr}(n, n+1; \mathbf{k})$. By duality, one may also write this imbedding as

$$\Sigma_0 \to \mathbf{Gr}(n, n+1; \mathbf{k}) \simeq \mathbf{Gr}(1, n+1; \mathbf{k}) = \mathbb{P}(\mathbf{k}^{n+1}),$$

where the composite isomorphism is also the Plucker imbedding. One may verify directly that, in this way, R induces a projective imbedding of Σ_0 as the affine ($\simeq \mathbf{A}^n$) piece

$$U_0 = \{[(p_0, p_1, \ldots, p_n)] : p_0 \neq 0\} \subset \mathbb{P}(\mathbf{k}^{n+1}).$$

For the sake of completeness, we describe the semistable and stable points.

PROPOSITION 4.8. *The null forms in* $\mathbf{A}^{n^2+n}$ *are those systems* (F, G) *where* F *is nilpotent. In contrast to the results of Section 2, there do not exist stable points for this action.*

Proof. Both statements repose on the equality

$$\mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})} = (\mathbf{k}[x_{ij}] \otimes_{\mathbf{k}} \mathbf{k}[x_1, \ldots, x_n])^{G\ell(n,\mathbf{k})};$$

see Lemma 4.3. In fact, the first statement now follows from the definitions (compare Proposition 2.9). As for the second statement, to say (F, G) is stable is to say there exists an invariant f with the action on $\mathbf{A}_f^{n^2+n}$ closed, and for which $(F, G) \in \mathbf{A}_f^{n^2+n}$. Now, since $\tilde{\Sigma}_0$ is Zariski open and since the restriction of $\phi_1 : \mathbf{A}^{n^2+n} \to \mathbf{A}^n$ to $\tilde{\Sigma}_0$ is surjective, $\tilde{\Sigma}_0$ intersects each fiber of ϕ_1 in a nonempty relatively open set, $\tilde{\Sigma}_0 \cap \phi^{-1}(p)$. Notice that $\tilde{\Sigma}_0 \cap \phi^{-1}(p)$ is just $\mathcal{O}(F_1, G_1)$, for (F_1, G_1) the completely controllable system with $\phi_1(F_1, G_1) = p$. Moreover, since the intersection $\tilde{\Sigma}_0 \cap \phi^{-1}(p) = \mathcal{O}(F_1, G_1)$ is proper, there is a system $(F_2, G_2) \in \mathscr{C}\ell(\mathcal{O}(F_1, G_1)) - \mathcal{O}(F_1, G_1)$; i.e., in each fiber of ϕ we may choose at least two distinct systems $(F_1, G_1), (F_2, G_2)$, for which $(F_2, G_2) \in \mathscr{C}\ell(\mathcal{O}(F_1, G_1))$. Now suppose we are given (F, G) and f as before, then since $f \in \mathbf{k}[x_{ij}]^{G\ell(n,\mathbf{k})}$ the entire fiber $\phi^{-1}(\phi(F, G))$ lies in $\mathbf{A}_f^{n^2+n}$. However, this excludes the possiblity that the action of $G\ell(n, \mathbf{k})$ on $\mathbf{A}_f^{n^2+n}$ be closed. Thus, $(\mathbf{A}^{n^2+n})_s = \varnothing$. Q.E.D.

Remark. In Section 5 we demonstrate the existence of canonical forms for arbitrary systems $(F, G) \in \mathbf{A}^{n^2+n}$, from which it follows that only finitely many orbits can lie in a given fiber of ϕ_1. By using the intertwining operator R, one may construct forms for systems $(F, G) \in \tilde{\Sigma}_0$ which are special cases of the canonical form known to exist for matrices in $M_*(n, n+1)$—or more generally in $M_*(n, \ell)$, $\ell \geqslant n$. Such forms constitute the local coordinates of the Grassmann variety and, in fact, these define the imbedding of Σ_0 into $\mathbf{P}(\mathbf{k}^{n+1})$ alluded to in the previous remark. Finally, this canonical form for completely controllable systems has been known for some time in another context, and is referred to as the "control canonical form," (see [18, Chap. 2]).

Now since the action α of $G\ell(n, \mathbf{k})$ on $\tilde{\Sigma}_0$ is free, the quotient $\tilde{\Sigma}_0 \to \Sigma_0$ has some rather spectacular properties. In fact, the existence of canonical forms implies that $\phi_1 : \tilde{\Sigma}_0 \to \Sigma_0$ is a "trivial" principal $G\ell(n, \mathbf{k})$-bundle. Explicitly, one may construct an algebraic section of ϕ_1 as follows. For $p \in \Sigma_0$, define F to be the companion matrix of $p(\lambda) = \lambda^n + \sum_{\ell=1}^{n} p_\ell \lambda^{n-\ell}$ and set $G = e_1$. Denoting this map by σ, notice that $\phi_1 \circ \sigma = id$.

LEMMA 4.9. *The map* $\alpha_1 : \Sigma_0 \times G\ell(n, \mathbf{k}) \to \tilde{\Sigma}_0$, *defined by* $\alpha_1(p, g) = \alpha(\sigma(p), g)$, *is a birational isomorphism.*

Proof. Clearly α_1 is rational. Since the action is free, α_1 is injective. For $(F, G) \in \tilde{\Sigma}_0$, $\sigma \circ \phi(F, G) = (F_1, G_1)$ consists of the rational canonical form,

F_1, of F and the vector $G = e_1$; i.e., (F_1, G_1) is the control canonical form of (F, G). Thus, the existence of the rational canonical form implies that α_1 is surjective. The proof that α_1^{-1} is rational involves calculating $\sigma \circ \phi_1$. Consider the classical canonical form for $R(F, G) \in M_*(n,n+1)$ given by $R(F,G)_1^{-1} R(F,G)$, where $R(F, G)_1$ is the invertible submatrix $(G \mid FG \mid \cdots \mid F^{n-1}G)$. Of course, the first column of $R(F, G)_1^{-1} R(F, G)$ is $G_1 = e_1$, while a straightforward calculation of the characteristic polynomial of $(FG \mid \cdots \mid F^nG)$ shows that this submatrix is the rational canonical form for F, i.e., $R(F, G)_1^{-1} R(F, G) = R(F_1, G_1)$. Using that R is an intertwining operator (see diagram (4.5) with $g = R(F, G)_1^{-1}$), it follows that α^{-1} may be calculated as

$$\alpha_1^{-1}: (F, G) \to (\phi_1(F_1, G_1), R(F, G)_1). \tag{4.6}$$

In particular, α_1^{-1} is rational. Q.E.D.

LEMMA 4.10. *The map* $\alpha_2 : \Sigma_0 \times G\ell(n, \mathbf{k}) \times \mathbf{A}^{pn} \to \tilde{\Sigma}$, *defined by* $\alpha_2(p, g, H) = (\alpha(\sigma(p), g), Hg^{-1})$ *is a birational isomorphism. Thus* α_2^{-1} *provides a canonical form for a system* $(F, G, H) \in \tilde{\Sigma}_0$, viz.,

$$\alpha_2^{-1}(F, G, H) = (\phi_1(F_1, G_1), R(F, G)_1, H \cdot R(F, G)_1). \tag{4.7}$$

Proof. A calculation, using the intertwining operator and diagram (4.5), shows that α_2^{-1} is indeed the inverse map for α_2. Now α_2 is rational and from (4.6) it follows that α_2^{-1} is rational as well. We are actually interested in the canonical form $(F_1, G_1, HR(F, G)_1)$ for a system (F, G, H). To show that this form is canonical it is clearly enough to show that α_2^{-1} is both constant on the orbits of $G\ell(n, \mathbf{k})$ and separates these orbits. To check the first statement it is easiest to introduce another intertwining operator, $R_1 : \tilde{\Sigma} \to M_*(n,n+1) \times \mathbf{A}^{pn}$, defined by $R_1(F, G, H) = (R(F, G), H)$, where $M_*(n, n+1) \times \mathbf{A}^{pn}$ admits the natural $G\ell(n, \mathbf{k})$-action, viz., $((M, H), g) \to (gM, Hg^{-1})$. That is to say, the diagram

$$\begin{array}{ccc} \tilde{\Sigma} & \xrightarrow{R_1} & M_*(n, n+1) \times \mathbf{A}^{pn} \\ \downarrow g & & \downarrow g \\ \tilde{\Sigma} & \xrightarrow{R_1} & M_*(n, n+1) \times \mathbf{A}^{pn} \end{array} \tag{4.8}$$

commutes for all $g \in G\ell(n, \mathbf{k})$. As is easily checked, α_2^{-1} is constant on the orbits and since the action on $\tilde{\Sigma}_0$ is free, the map α_2 separates these orbits. Thus the form given by (4.7) is canonical. Q.E.D.

By Mumford's theorem (see [23 Theorem 1.1] a (universal) categorical quotient of $\tilde{\Sigma}$ by $G\ell(n, \mathbf{k})$ exists. One may show directly that any invariant, f, on $\tilde{\Sigma}$ descends to a unique map $f_1 : \Sigma_0 \times \mathbf{A}^{pn} \to \mathbf{A}^1$, satisfying $f_1 \circ \phi_2 = f$, where ϕ_2 is α_2^{-1} followed by the natural projection. Thus, $\phi_2 : \tilde{\Sigma} \to \tilde{\tilde{\Sigma}} = \Sigma_0 \times \mathbf{A}^{pn}$ is a categorical quotient. Since the action on $\tilde{\Sigma}$ is free, from [23, Amplification 1.3], it now follows:

THEOREM 4.11. *$\phi_2 : \tilde{\Sigma} \to \Sigma$ is a (universal) geometric quotient.*

Remarks. 1. Several authors have noticed, in case $p = 1$, that the orbit set $\tilde{\Sigma}/G\ell(n, \mathbf{k})$ is in bijection with $\Sigma_0 \times \mathbf{k}^n = \mathbf{k}^{2n}$. In fact, for arbitrary p, Kalman's techniques (see [16, 17]) work very well and yield the quotient set $\mathbf{k}^{2pn}$.

2. The proper generalization of Theorem 4.11, for $m > 1$, seems to stem from viewing Σ as a vector bundle with base space Σ_0 (see [4, 5]). In fact, since the action on $\tilde{\Sigma}_0$ is free, $\tilde{\Sigma}_0 \to \Sigma_0$ carries the structure of a principal $G\ell(n, \mathbf{k})$-bundle by [23, Proposition 0.9]. A careful reading of the proof of Lemma 4.10 shows that Σ may be regarded, via the intertwining map R_1, as the $\mathbf{A}^{pn}$ bundle associated to $\tilde{\Sigma}_0 \to \Sigma_0$.

Since the quotient $\phi_2 : \tilde{\Sigma} \to \Sigma$ is geometric, one may describe open sets in Σ as exactly those sets $V \subset \Sigma$ for which $\phi_2^{-1}(V)$ is an invariant open subset of $\tilde{\Sigma}$. Kalman has asked (see [17]) the following question: Which systems-theoretic properties are "generic," in the sense that the systems with such a property form a Zariski open subset of Σ? Since a systems-theoretic property either holds or does not hold on a given orbit, a classification of generic properties reduces to a classification of those properties P for which the set $\mathscr{P} = \{(F, G, H) \in \tilde{\Sigma} : (F, G, H) \text{ has } P\}$ is defined by polynomial inequalities. For example, another important system-theoretic property is complete observability viz., a system (F, G, H) is completely observable provided each state is detectable in terms of the various outputs. Formally,

DEFINITION 4.12. *(F, G, H) is completely observable if and only if the matrix* $(H^t \mid F^tH^t \mid \cdots \mid (F^t)^{n-1}H^t)$ has rank n. The following assertion is clear.

PROPOSITION 4.13. *Complete observability is a generic property for completely controllable, scalar input linear dynamical systems.*

COROLLARY 4.14 (Hermann [12, Chap. 8, Sect. 4]). *The orbits of completely controllbable and completely observable, scalar input–output systems form a Zariski open subspace of* $\mathbf{A}^{2n+n^2}$.

Notice that there is a natural duality at play, viz., (F, G, H) is completely observable if and only if (F^t, H^t, G^t) is completely controllable. Thus let $\tilde{\Sigma}^*$ be the set of all LDS's which are completely observable and have a scalar output. As before, $\tilde{\Sigma}_0{}^* = \{(F, 0, H) \in \tilde{\Sigma}^*\}$. By duality, one may obtain canonical forms for systems $(F, G, H) \in \tilde{\Sigma}^*$. Moreover, THEOREM 4.15. *A geometric quotient Σ^* of $\tilde{\Sigma}^*$ by $G\ell(n, \mathbf{k})$ exists. $\Sigma^* \simeq \Sigma_0^* \times \mathbf{A}^{nm}$, where $\Sigma_0^* \simeq \mathbf{A}^n$ is the geometric quotient of $\tilde{\Sigma}_0^*$ by $G\ell(n, \mathbf{k})$.*

COROLLARY 4.16. *Complete controllability is a generic property for systems in Σ^*.*

Kalman [17] also asked whether the following property is generic. Call a system cyclic if there is a column in G which is a cyclic vector for F. In $\tilde{\Sigma}^*$, these are the systems which may be controlled through a single input terminal.

COROLLARY 4.17. *Cyclicity is a generic property for systems in Σ^*.*

Finally, we indicate how the results of this section may be extended to two other classes of linear dynamical systems. First, if $\mathbf{k} = \mathbf{C}$ then we may define a continuous-time LDS as before, where the equations of evolution should be redefined as

$$\dot{x}(t) = Fx(t) + Gu(t), \tag{4.9}$$

$$y(t) = Hx(t) + Ku(t). \tag{4.10}$$

For continuous-time systems, algebraic equivalence (mod $G\ell(n, \mathbf{k})$) is not implied, in general, by systems-theoretic equivalence. However, this is the case when (F, G, H, K) is taken to be both completely controllable and completely observable (see [12, Chap. 1, Theorem 5.2]). Setting $K = 0$, temporarily, Proposition 4.13 (or Corollary 4.16) implies the existence of a quasi-projective quotient for such systems. Moreover, the appropriate open subspace of Σ (or of Σ^*) carries the structure of a complex analytic manifold. Thus a moduli space for such continuous-time systems exists within the category of analytic manifolds. Second, let $\tilde{S}$ denote the set of all scalar input LDS's (F, G, H, K) which are completely controllable. Then a geometric quotient exists for the action of $G\ell(n, \mathbf{k})$ on $\tilde{S}$, viz., $\phi_3 : \tilde{S} \to S = \Sigma \times \mathbf{A}^p \simeq \mathbf{A}^{p(n+1)+n}$. The map ϕ_3 assigns to a system (F, G, H, K) the canonical form $(\phi_2(F, G, H), K)$, see Lemma 4.10 and the second remark following Definition 4.1. Dually, if $\tilde{S}^*$ denotes the set of completely observable, scalar output systems, one has a geometric quotient

$$\phi_3^* : \tilde{S}^* \to S^* \simeq \Sigma^* \times \mathbf{A}^m \simeq \mathbf{A}^{m(n+1)+n}.$$

Similar remarks apply to the moduli of continuous-time systems. There are generalizations of the continuous-time construction, for arbitrary complete fields, topologized by a nontrivial absolute value (see [5]). Also, one may construct the various moduli spaces for arbitrary input–output systems (see [4, 5]).

One may also consider a more general type of scalar input system. In the next section, we construct canonical forms for arbitrary systems in $\mathbf{A}^{n^2+n}$ and, using the Main Theorem of Section 3, we derive rational criteria for deciding whether two systems are equivalent.

5. Canonical Forms and Decidability

Criteria for Scalar Input Systems

In this section we are concerned with scalar input systems (i.e., pairs in $M(n, \mathbf{k}) \times \mathbf{k}^n$). As in Theorem 3.6, our goal is to derive algebraic criteria for deciding the equivalence problem for the relation of state-space equivalence (see Definition 4.1). That the proof of Theorem 3.6 was possible was due largely to the existence of the Jordan canonical form. Thus the first order of business here is to obtain a canonical form for a scalar input system (F, G). This form differs from the "control canonical form" considered for completely controllable systems not only in the generality of systems considered, but also in the type of form employed, viz., the Jordan form versus the rational form. The idea is rather simple. (F, G) is equivalent to a system (F_J, G_1) where F_J is the Jordan form of F. One can then bring (F_J, G_1) to a canonical form by working on G_1 with elements of $G\ell(n, \mathbf{k})$ centralizing F_J, that is, the Jordan form of F is going to be the first term of the canonical form of the pair (F, G).

Let $Z(F)$ be the centralizer of F in $M(n, \mathbf{k})$. Write $F = F_{\lambda_1} \oplus \cdots \oplus F_{\lambda_s}$ where $F_{\lambda_i} = J(\lambda_i ; n_{ir_i}) \oplus \cdots \oplus J(\lambda_i ; n_{i1})$, $n_{ir_i} \geqslant \cdots \geqslant n_{i1}$ and $\lambda_i \neq \lambda_j$ if $i \neq j$. Set $m_i = \Sigma_j n_{ij}$, the multiplicity of λ_i. Then $Z(F) = \oplus_{i=1}^s Z(F_{\lambda_i})$; $Z(F_{\lambda_i}) \subseteq M(m_i, \mathbf{k})$. Now $Z(F_{\lambda_i})$ can be described in terms of its r_i^2 subblocks. The (k, ℓ)th subblock is n_{ik} by $n_{i\ell}$, and is of the form

$$\left|\begin{matrix} \omega_1 & \omega_2 & \cdots & \omega_{n_{i\ell}} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \omega_2 \\ & & & \omega_1 \\ & & & 0 \\ & & & \vdots \\ 0 & & \cdots & 0 \end{matrix}\right| \qquad \text{if} \quad k \leqslant \ell \; (n_{ik} \geqslant n_{i\ell}) \tag{5.1}$$

or

$$\left|\begin{matrix} 0 & \cdots & 0 & \omega_1 & \omega_2 & \cdots & \omega_{n_{ik}} \\ \vdots & & & \ddots & \ddots & & \vdots \\ & & & & \ddots & \ddots & \omega_2 \\ 0 & & \cdots & & & 0 & \omega_1 \end{matrix}\right| \qquad \text{if} \quad k \geqslant \ell \; (n_{ik} \leqslant n_{i\ell}), \tag{5.2}$$

where the ω's belong to $\mathbf{k}$ and can vary freely from block to block.

Let G_p be the group of all $p \times p$ matrices of the form

$$\begin{vmatrix} \omega_1 & \omega_2 & \cdots & \omega_p \\ & \ddots & \ddots & \vdots \\ & & \ddots & \omega_2 \\ & & & \omega_1 \end{vmatrix}, \qquad \omega_1 \neq 0. \tag{5.3}$$

As the next step in obtaining a canonical form for the pair (F, G), we obtain $\mathbf{k}^p/G_p$. Let e_i be the $p \times 1$ matrix with 1 in the ith row and zeroes elsewhere.

PROPOSITION 5.1. $\mathbf{k}^p/G_p = \{0, e_1, ..., e_p\}$.

Proof. Go by induction on p. First of all, G_p stabilizes $\langle e_1 \rangle$, and hence induces an action on $\mathbf{k}^{p-1} = \mathbf{k}^p/\langle e_1 \rangle$. The image of G_p in $G\ell(p-1, \mathbf{k})$ under this action is G_{p-1}. Thus, given v in $\mathbf{k}^p$, by the induction assumption it can be brought by the action of G_p to the form xe_1 or $xe_1 + e_i$ $(i > 1)$ for some scalar x. The first type is evidently in the orbit of 0 or e_1 depending on whether x is zero or not. If $x \neq 0$, the second type can be transformed to e_i by an element of G_p for which $\omega_1 = 1, \omega_2 = \cdots = \omega_{i-1} = 0, \omega_i = -x$.

To see that $0, e_1, ..., e_p$ lie in distinct orbits we calculate the dimensions of their orbits. Now $\dim \mathcal{O}(0) = 0$. $\mathcal{O}(e_i)$ is the image of the map $\alpha_i : G_p \to \mathbf{k}^p$ given by $g \to g(e_i)$, hence the dimension of $\mathcal{O}(e_i)$ is the rank of the differential of this map at I. $d\alpha_i(I): L(G_p) \to \mathbf{k}^p$ by $h \to h(e_i)$ and $L(G_p)$, the Lie algebra of G_p, is the set of all matrices of the form (5.3) where even ω_1 is allowed to be zero. If g is given by (5.3) then $g(v) = d\alpha_i(I)(g) = \omega_i e_1 + \omega_{i-1} e_2 + \cdots + \omega_1 e_i$. Thus $\dim \mathcal{O}(e_i) = \operatorname{rank} d\alpha_i(I) = i$.

DEFINITION. If (F, G) belongs to $M(n, \mathbf{k}) \times \mathbf{k}^n$, we define $\operatorname{gr}(F, G)$, the grade of G with respect to F, to be the dimension of the F-submodule of $\mathbf{k}^n$ generated by G. Equivalently, it is the rank of the matrix whose columns are $G, FG, ..., F^{n-1}G$.

Remarks. 1. $\operatorname{gr}(F, G)$ is an equivalence invariant of the system (F, G).

2. $\operatorname{gr}(F, G)$ has several useful properties. Write $\mathbf{k}^n = \mathbf{k}^n_{\lambda_1} \oplus \cdots \oplus \mathbf{k}^n_{\lambda_s}$, $G = \Sigma G_i$, where $\mathbf{k}^n_{\lambda_i}$ is the λ_i-eigenspace of $\mathbf{k}^n$ with respect to F, G_i is in $\mathbf{k}^n_{\lambda_i}$, and $\lambda_i \neq \lambda_j$ if $i \neq j$. Then $\operatorname{gr}(F, G) = \Sigma \operatorname{gr}(F, G_i)$. On the other hand, if $\mathbf{k}^n_{\lambda_i} = \mathbf{k}^n_{i1} \oplus \cdots \oplus \mathbf{k}^n_{ir_i}$ is a decomposition corresponding to Jordan λ_i-blocks of F, and if $G_i = G_{i1} \oplus \cdots \oplus G_{ir_i}$, then $\operatorname{gr}(F, G_i) = \max\{\operatorname{gr}(F, G_{ij})\}_j$.

3. The group G_p in Proposition 5.1 is the centralizer of $J(\lambda; p)$ acting on $\mathbf{k}^p$. It is easily checked that $\mathcal{O}(e_i)$ is the set of all vectors of grade i with respect to $J(\lambda; p)$.

To avoid a proliferation of indices, we look at the canonical form problem one eigenvalue at a time. Suppose we have a system (F, G) in $M(n, \mathbf{k}) \times \mathbf{k}^n$, such that $F = J(\lambda; n_r) \oplus \cdots \oplus J(\lambda; n_1)$ where $n_r \geqslant \cdots \geqslant n_1$. Identify $\mathbf{k}^n$ with $\mathbf{k}^{n_r} \oplus \cdots \oplus \mathbf{k}^{n_1}$ in an obvious way corresponding to the blocks in F, and write $G = G_r + \cdots + G_1$ where G_j belongs to $\mathbf{k}^{n_j}$. We let $e_{j1}, \dots, e_{jn_j}$ be the standard basis vectors in $\mathbf{k}^{n_j}$. By Proposition 5.1, for each j we can assume G_j is either 0 or e_{jk} for some k depending on j. Now $k = \mathrm{gr}(F, e_{jk})$. Recalling that $\mathrm{gr}(F, G) = \max\{\mathrm{gr}(F, G_j)\}_j$, if $G \neq 0$, pick the smallest number j for which $\mathrm{gr}(F, G) = \mathrm{gr}(F, G_j)$. Then we claim that $(F, G) \sim (F, G_j)$. For notice that if $k < j$ one has $\mathrm{gr}(F, G_k) < \mathrm{gr}(F, G_j)$, and if $k > j$ then $\mathrm{gr}(F, G_k) \leqslant \mathrm{gr}(F, G_j)$. Also according to (5.1) and (5.2),

$$\begin{aligned}&\text{the space } Z(F)(e_{jk}) \text{ is spanned by all } e_{i\ell} \text{ where either } \ell \leqslant k \text{ and } i\\ &\text{is such that } n_i \geqslant n_j\text{, or } \ell < k \text{ and } i \text{ is such that } n_j > n_i .\end{aligned} \tag{5.4}$$

Thus there is a nonsingular g in $Z(F)$ which takes G_j to G. Such a g transforms (F, G_j) to (F, G).

To clarify the argument for this claim further, we feel it is worthwhile to include some examples. So suppose $F = J(\lambda; 3) \oplus J(\lambda; 2)$. Then $Z(F)$ is given by all

$$\begin{vmatrix} x_0 & x_1 & x_2 & y_0 & y_1 \\ 0 & x_0 & x_1 & 0 & y_0 \\ 0 & 0 & x_0 & 0 & 0 \\ 0 & z_0 & z_1 & t_0 & t_1 \\ 0 & 0 & z_0 & 0 & t_0 \end{vmatrix}, \qquad x_i, y_i, z_i, t_i \in \mathbf{k}.$$

The corresponding basis vectors of $\mathbf{k}^5$ are written $e_{21}, e_{22}, e_{23}, e_{11}, e_{12}$ in our notation above. Then $G = e_{22} + e_{12}$ is of grade 2 and $(F, G) \sim (F, e_{12})$ by

$$g = \begin{vmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

Or, if $G = e_{23} + e_{12}$, G is of grade 3 and $(F, G) \sim (F, e_{23})$ by

$$g = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{vmatrix}.$$

Thus, in the case that F has only one eigenvalue, we have shown that any system (F, G) is equivalent to a system (F_J, G_1) where F_J is the Jordan form of G, and G_1 is either 0 or else it has one nonzero component which is a 1. It remains to show that if (F_J, G_2) is another such pair which is equivalent to (F, G) that $G_1 = G_2$. Note first that $(F_J, G_1) \sim (F_J, G_2)$, hence by Remark 1. above, $\mathrm{gr}(F_J, G_1) = \mathrm{gr}(F_J, G_2)$. Also, there is a nonsingular g in $Z(F_J)$ such that $gG_1 = G_2$. But since G_1 and G_2 have the same grade, Eqs. (5.1) and (5.2) force G_1 and G_2 to have their ones in the same block. (Nothing in $Z(F_J)$ can move G_1 to something of the same grade having its one in a smaller-sized block.) Having their ones in the same block and being the same grade forces $G_1 = G_2$. We have thus proven most of

THEOREM 5.2. (Canonical form for scalar input systems). *Every scalar input system (F, G) is equivalent to exactly one system (F_J, G_J) where F_J is the Jordan form of G, and G_J has at most one nonzero component, which is a one, in each F_J-eigenspace of $\mathbf{k}^n$.*

Proof. It is clear that every system (F, G) has at least one such form. It remains to prove the uniqueness in the several eigenvalue case. But this follows directly from the one-eigenvalue case in view of the fact that elements of $G\ell(n, \mathbf{k})$ centralizing F_J also stabilize the F_J-eigenspaces in $\mathbf{k}^n$.

Remarks. 1. Kalman [15] and Popov [25] have given generalizations of the control canonical form, valid for systems (F, G) with some restrictions on the rank of G and the rank of $R(F, G)$(see Sect. 4). The forms introduced in [15] are actually canonical forms for the action of a larger group than $G\ell(n, \mathbf{k})$, taking into account more internal symmetries and the introduction of feedback.

2. The forms obtained here (as well as those obtained in [15, 25]) apply to systems whose output is just the change of state, i.e. $H = 0$. The obstructions to putting (F, G, H) into canonical form lie in the subgroup of $G\ell(n, \mathbf{k})$ stabilizing (F, G). As we have seen in Section 4 this subgroup is trivial if and only if (F, G) is completely controllable.

THEOREM 5.3 (Decidability criteria for the equivalence of scalar input systems). *Suppose (F, G) and (D, E) are scalar input systems, i.e., elements of $M(n, \mathbf{k}) \times \mathbf{k}^n$) over an algebraically closed field of characteristic zero. Then (F, G) is equivalent to (D, E) if and only if*

(i) *F and D have the same characteristic polynomial,*

(ii) $\mathrm{rank}(F \otimes 1 - 1 \otimes F)^k = \mathrm{rank}(D \otimes 1 - 1 \otimes D)^k = \mathrm{rank}(F \otimes 1 - 1 \otimes D)^k$ *for all even $k \leqslant 2n - 2$,*

(iii) $\mathrm{gr}(F \oplus F, G \oplus G) = \mathrm{gr}(D \oplus D, E \oplus E) = \mathrm{gr}(F \oplus D, G \oplus E)$,

(iv) $\dim \mathcal{O}(F \oplus F, G \oplus G) = \dim \mathcal{O}(D \oplus D, E \oplus E)$
$= \dim \mathcal{O}(F \oplus D, G \oplus E)$.

Proof. The necessity is direct. For the sufficiency, since (i)–(iv) are invariant under equivalence, we can assume (F, G) and (D, E) to be in their canonical form as in Theorem 5.2. Then Theorem 3.6 allows us to assume that $F = D$.

So suppose that $G = G_1 + \cdots + G_s$, $E = E_1 + \cdots + E_s$ are decompositions of G and E whose components G_i, E_i belong to $\mathbf{k}^n_{\lambda_i}$, the λ_i-eigenspace of $F = D$ on $\mathbf{k}^n$, for distinct eigenvalues. Let $g_i = \operatorname{gr}(F, G_i)$, $h_i = \operatorname{gr}(F, E_i)$. By Remark 2. after Proposition 5.1, one has

$$\begin{aligned}\operatorname{gr}(F \oplus F, G \oplus G) &= \sum_i \operatorname{gr}(F \oplus F, G_i \oplus G_i)\\ &= \sum_i g_i, \operatorname{gr}(D \oplus D, E \oplus E) = \sum h_i,\end{aligned}$$

and $\operatorname{gr}(F \oplus F, G \oplus E) = \sum_i \operatorname{gr}(F \oplus F, E_i \oplus G_i) = \sum_i \max(g_i, h_i)$. Condition (iii) of the theorem and the last three equations force $\operatorname{gr}(F, G_i) = \operatorname{gr}(F, E_i)$ for all i.

Each G_i, E_i is either zero, or else it has one nonzero entry which is a one. Let $n_i(n_i')$ be the size of the λ_i-block of F to which $G_i(E_i)$ belongs (we block up $\mathbf{k}^n$ according to the way F is blocked, and recall that F is in Jordan form). Then, repeating arguments used to set up the canonical form in the one eigenvalue case, since G_i and E_i have the same grade, the canonical form of $(F \oplus F, G \oplus G)$ $((F \oplus F, E \oplus E))$ is $(F \oplus F, G \oplus 0)((F \oplus F, E \oplus 0))$, while the canonical form of $(F \oplus F, G \oplus E)$ is $(F \oplus F, B)$ where $B = \sum B_i$, $\operatorname{gr}(F \oplus F, B_i) = \operatorname{gr}(F, G_i) = \operatorname{gr}(F, E_i)$, $B_i \in \mathbf{k}^{2n}_{\lambda_i}$, and B_i is either zero or has its one in a λ_i-block of size $\min(n_i, n_i')$.

Condition (iv) here together with (4.4) implies

$$\begin{aligned}\dim Z(F \oplus F)(F \oplus 0) &= \dim Z(F \oplus F)(E \oplus 0)\\ &= \dim Z(F \oplus F)(G \oplus E) = \dim Z(F \oplus F)(B).\end{aligned} \tag{5.5}$$

The centralizer of a matrix is the diagonal sum of the centralizers of the various eigenvalue components (i.e. the diagonal sum of all blocks for a fixed eigenvalue). Since G_i, E_i, B_i have the same grade, (5.4) implies that $\dim Z(F \oplus F)(B_i) = \max(\dim Z(F \oplus F)(G_i \oplus 0), \dim Z(F \oplus F)(E_i \oplus 0))$. Now (5.5), the equation of the last line, and the identities $\dim Z(F \oplus F)(G \oplus 0) = \sum_i \dim Z(F \oplus F)(G_i \oplus 0)$, $\dim Z(F \oplus F)(E \oplus 0) = \sum_i \dim Z(F \oplus F)(E_i \oplus 0)$, and $\dim Z(F \oplus F)(B) = \sum_i \dim Z(F \oplus F)(B_i)$, imply that $\dim Z(F \oplus F)(G_i \oplus 0) = \dim Z(F \oplus F)(E_i \oplus 0)$ for all i. Since E_i and G_i already are known to have the same grade, (5.4) and the last equation force them to have their ones in the same size λ_i-block, thus $G = E$.

6. Conjugacy in the Classical Groups and Their Lie Algebras

Let $\mathbf{k}$ be an algebraically closed field of characteristic zero. The similarity problem can be viewed as a conjugacy problem in either the special linear group or its Lie algebra. (Two elements of a Lie algebra are called conjugate if there is an inner automorphism of the Lie algebra carrying the one element onto the other.) Since scalar matrices are central, $SL(n, \mathbf{k})$ has the same conjugacy classes in $M(n, \mathbf{k})$ as $GL(n, \mathbf{k})$. $SL(n, \mathbf{k})$ is the classical group of type A_{n-1}, and its Lie algebra, the trace-zero matrices, is the simple Lie algebra of type A_{n-1}. It is well-known that every inner automorphism of this Lie algebra is induced by conjugation with respect to some element of $SL(n, \mathbf{k})$. Thus, if X is either the special linear group or its Lie algebra, Theorem 3.6 gives decidability criteria for the conjugacy problem in X. Due to a result of Freudenthal and Theorem 3.6, we can give a similar criteria for the conjugacy problem in the other classical groups and their Lie algebras.

Let $(\ , \ )$ be a nondegenerate symmetric or skew-symmetric form on a finite-dimensional space V. One has associated to this form a group G consisting of all T in $GL(V)$ such that $(T(v_1), T(v_2)) = (v_1, v_2)$ for all v_i in V. When the form is symmetric and $\dim(V) = 2n + 1$ ($\dim(V) = 2n$), G is the orthogonal group of type $B_n(D_n)$. When the form is skew-symmetric, $\dim(V) = 2n$ and G is the symplectic group of type C_n. Also associated to this form is a Lie algebra L consisting of all T in $\operatorname{End}(V)$ satisfying $(T(v_1), v_2) = -(v_1, T(v_2))$ for all v_i in V. L is the Lie algebra of the corresponding group G. Furthermore, the inner automorphisms of L are known to be induced by conjugation with respect to elements of G, just as was the case for the special linear group. The description of all the classical groups and Lie algebras above is via an identification with their first fundamental representations. We give Freudenthal's [29] result next, including its proof (with a slight addition) for the sake of completeness.

THEOREM 6.1 (Freudenthal). *Let $(\ , \ )$ be a nondegenerate symmetric or skew-symmetric form on a finite-dimensional space V over an algebraically closed field of characteristic not 2. Let G be the subgroup of $GL(V)$ preserving the form, and let L be the Lie subalgebra of* $\operatorname{End}(V)$ *of skew elements with respect to the form. Let both A and B be elements of either G or L. Then $DAD^{-1} = B$ for some D in $GL(V)$ if and only if $DAD^{-1} = B$ for some D in G.*

Proof. For any T in $\operatorname{End}(V)$ we denote by T^t the transpose of T with respect to the form. The sufficiency is to be established. So suppose $DAD^{-1} = B$ for some D in $GL(V)$. Taking transposes, one obtains either $-(D^t)^{-1} AD^t = -B$ or $(D^t)^{-1} A^{-1} D^t = B^{-1}$ depending, respectively, on whether A and B belong to L or G. (A in L means $A^t = -A$, while A in G means $A^t = A^{-1}$.) In either case one obtains $(D^t)^{-1} AD^t = B$. Now set $S = D^{-1}(D^t)^{-1}$. Notice that S is in $GL(V)$ and $S^t = S$. It is easily checked that $S^tA = AS$, or equivalently, that

$AS = SA$. Hence for any polynomial $p(S)$ one has that $p(S)\,A = Ap(S)$. Pick a polynomial $r(S) = T$ with $T^2 = S$. Then (i) $T^t = T$, (ii) $T^2 = S = D^{-1}(D^t)^{-1}$, so $TD^t = T^{-1}D^{-1}$, (iii) $TA = AT$, and (iv) T is nonsingular because S is nonsingular.
Hence

$$(TD^t)^t\, A(TD^t) = DT^tAT\, D^t = DAT^2\, D^t = DA\, D^{-1}(D^t)^{-1}\, D^t = DA\, D^{-1} = B.$$

Also

$$(T\, D^t)^{-1} = (T^{-1}\,{}^{-1})^{-1} = DT = (T^t\, D^t)^t = (T\, D^t).$$

That is, TD^t belongs to G.

Corollary 6.2. (Decidability criteria for conjugacy in classical groups and Lie algebras). *Let X be either a classical group or its Lie algebra, and let α be its first fundamental representation. Then x and y in X are conjugate if and only if $\alpha(x)$ and $\alpha(y)$ are similar. Hence, if the characteristic of* $\mathbf{k}$ *is zero, x is conjugate to y if and only if*

(i) *$\alpha(x)$ and $\alpha(y)$ have the same characteristic polynomial, and*

(ii) *for all even $k \leqslant 2\dim(\alpha) - 2$*

$$\begin{aligned}\operatorname{rank}(\alpha(x) \otimes 1 - 1 \otimes \alpha(x))^k &= \operatorname{rank}(\alpha(y) \otimes 1 - 1 \otimes \alpha(y))^k \\ &= \operatorname{rank}(\alpha(x) \otimes 1 - 1 \otimes \alpha(y))^k.\end{aligned}$$

Of course it would be interesting to see if this result can be extended to all simple algebraic groups and their Lie algebras.

References

1. A. C. Aitken and H. W. Turnbull, "An Introduction to the Theory of Canonical Matrices," Dover, New York, 1931.
2. M. F. Atiyah and I. G. MacDonald, "Introduction to Commutative Algebra," Addison–Wesley, Reading, Mass., 1969.
3. A. Borel, "Linear Algebraic Groups," Benjamin, New York, 1969.
4. C. I. Byrnes and N. E. Hurt, The moduli of linear dynamical systems. I. Discrete time systems (to appear).
5. C. I. Byrnes and N. E. Hurt, The moduli of linear dynamical systems. II. Continuous time systems and design (to appear).
6. L. E. Dickson, Canonical forms of a linear homogeneous transformation in an arbitrary realm of rationality, *Amer. J. Math.* **24** (1902), 101–108.
7. L. E. Dickson, "Modern Algebraic Theories," Sanborn, Chicago, 1926.
8. J. Dieudonné and J. Carrell, Invariant theory, old and new, *Advances in Math.* **4** (1970), 1–80.

9. J. Fogarty, "Invariant Theory," Benjamin, New York, 1969.
10. G. Frobenius, Theorie der linearen formen mit ganzen coefficienten, *J. fur Math.* **86** (1879), 146–208.
11. M. Gauger, Elementary Divisors of Induced Linear Transformations on Tensors via the Representation Theory of sl_2, *Linear and Multilinear Algebra,* to appear.
12. R. Hermann, "Linear Systems Theory and Introductory Algebraic Geometry," Vol. 8, Interdisciplinary Mathematics, Brookline, Mass., 1974.
13. C. G. J. Jacobi, Uber eine elementary Transformation eines in Bezug auf jedes von zwei Variablen-Systemen linearen und homogenen Ausdrucken, *J. fur Math.* **53** (1857), 265–270; or *Werke*, **3**, 484–490.
14. C. Jordan, "Traité des Substitutions et des Équations Algébriques," Paris, 1870.
15. R. E. Kalman, Kronecker invariants and feedback,(Conference on Ordinary Differential Equations, June 14–23, 1971.
16. R. E. Kalman, Global structure of classes of linear dynamical systems, *in* "Lectures, NATO Advanced Study Institute on Geometric and Algebraic Methods for Non-Linear Systems, London, August–September, 1971."
17. R. E. Kalman, Algebraic geometric description of the class of linear systems of constant dimension, *in* "Proceedings of the Eighth Annual Princeton Conference on Information Sciences and Systems, Princeton, N.J., March 1974," to appear.
18. R. E. Kalman, P. L. Falb, and M. A. Arbib, "Topics in Mathematical Systems Theory," McGraw–Hill, New York, 1969.
19. G. Kowalewski, *Berlin Verh. sachs Akad. Leipzig* **68** (1916), 325–335.
20. S. Lattès, *Ann. Fac. Sci. Univ. Toulouse* **28** (1914), 1–84.
21. C. C. MacDuffee, "The Theory of Matrices," Chelsea, New York, 1956.
22. D. Mumford, "Introduction to Algebraic Geometry," Harvard Univ. Press, Cambridge, Mass., 1965.
23. D. Mumford, "Geometric Invariant Theory," Ergeb. Math. Bd. 34, Springer–Verlag, Berlin/New York, 1965.
24. D. Mumford and K. Suominen, Introduction to the theory of moduli, *in* "Proceedings of the Fifth Summer School in Math., Oslo, 1970," pp. 171–222.
25. V. M. Popov, Invariant description of linear time-invariant controllable systems, *SIAM J. Control* (1972).
26. W. E. Roth, On direct product matrices, *Bull. Amer. Math. Soc.* **40** (1934), 461–468.
27. E. Weyr, *C. R. Acad. Sci. Paris* **100** (1885), 966–969.
28. E. Weyr, *Mh. Math. Phys.* **1** (1890), 163–236.
29. H. Freudenthal, Elementarteilertheorie der Komplexen orthagnalen und symplektischen Gruppen, *Indag. Math.* **14** (1952), 99–101.